

REVISED
VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : A61K 38/00, C07K 1/00, 16/00, C12Q 1/68, C12P 19/34, C07H 21/02, 21/04		(11) International Publication Number: WO 99/42118
A2		(43) International Publication Date: 26 August 1999 (26.08.99)
(21) International Application Number: PCT US99 03265		(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).
(22) International Filing Date: 17 February 1999 (17.02.99)		
(30) Priority Data: 09/024,753 18 February 1998 (18.02.98) US 09/072,596 5 May 1998 (05.05.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(71) Applicant: CORIXA CORPORATION (US US); 1424 Colum- bia Street, Seattle, WA 98104 (US).		
(72) Inventors: REED, Steven, G.; 2843 122nd Place N.E., Belle- vue, WA 98005 (US). SKEIKY, Yasar, A.; 8327 25th Street, Seattle, WA 98107 (US). DILLON, David, C., 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAM- POS-NETO, Antonio; 9308 Midship Court N.E., Bain- bridge Island, WA 98110 (US). HOUGHTON, Ray- mond; 2636 242nd Place S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 36th Avenue S.W., Seattle, WA 98126 (US). HENDRICKSON, Ronald, C.; 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).		

Published

With declaration under Article 17(2)(a)

*Without abstract, title not checked by the International
Searching Authority*

(54) Title: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT

AE	United Arab Emirates	ES	Spain	IS	Iceland	SE	Sweden
AM	Armenia	FI	Finland	IT	Italy	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AX	Åland Islands	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Moldova	TG	Togo
BB	Barbados	GR	Greece	MG	Madagascar	TH	Thailand
BE	Belgium	HN	Honduras	MK	Macedonia	TM	Turkmenistan
BF	Burkina Faso	HR	Croatia	ML	Mali	TR	Turkey
BG	Bulgaria	IE	Ireland	MT	Malta	TT	Trinidad and Tobago
BI	Burundi	IL	Israel	MX	Mexico	UA	Ukraine
BR	Brazil	IN	India	MY	Malaysia	UG	Uganda
BS	Bahamas	IS	Costa Rica	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MY	Myanmar	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NP	Nepal	VN	Vietnam
CG	Congo	KR	South Korea	NI	Nicaragua	YE	Yemen
CH	Switzerland	KW	Kuwait	NO	Norway	ZW	Zimbabwe
CI	Ivory Coast			NZ	New Zealand		
CM	Cameroun						

REVISED
VERSION

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT
(PCT Article 17(2)(a) and Rule 39)

Applicant's or agent's file reference 9532-023-228	IMPORTANT DECLARATION	Date of mailing (day/month/year) 22 JUNE 1999
International application No. PCT/US99/03265	International filing date (day/month/year) 17 FEBRUARY 1999	(Earliest) Priority Date (day/month/year) 18 FEBRUARY 1998
International Patent Classification (IPC) or both national classification and IPC Please See Continuation Sheet.		
Applicant CORIXA CORPORATION		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below.

1. ☐ The subject matter of the international application relates to:
- a. ☐ scientific theories.
 - b. ☐ mathematical theories.
 - c. ☐ plant varieties.
 - d. ☐ animal varieties.
 - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
 - f. ☐ schemes, rules or methods of doing business.
 - g. ☐ schemes, rules or methods of performing purely mental acts.
 - h. ☐ schemes, rules or methods of playing games.
 - i. ☐ methods for treatment of the human body by surgery or therapy.
 - j. ☐ methods for treatment of the animal body by surgery or therapy.
 - k. ☐ diagnostic methods practiced on the human or animal body.
 - l. ☐ mere presentations of information.
 - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
- ☐ the description
 - ☐ the claims
 - ☐ the drawings
3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the prescribed requirements prevents a meaningful search from being carried out:
- ☒ it does not comply with the prescribed standard
 - ☐ it is not in the prescribed machine-readable form
4. Other reasons:

**DECLARATION OF NON-ESTABLISHMENT OF
INTERNATIONAL SEARCH REPORT**

International application No.
PCT/US99/03265

The International Patent Classification (IPC) or National Classification and IPC are as listed below:

IPC(5): A61K 38/00; C07K 1/00; C07K 16/00; C12Q 1/68; C12P 19/34; C07H 21/02, 21/04
US Cl. 530/300, 350, 387.1; 435/6, 91.1, 91.2; 536/23.1, 24.3, 24.31, 24.32, 24.33

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann, in
5 *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis. The present invention fulfills this need and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods for diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen
15 that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);
- (d) Thr-Lys-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID

- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 122) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 184-188, 194-196, 198, 210-220, 232, 234, 235, 237-242, 248-251, 256-271, 287, 288, 290-293 and 298-337, ,
5 the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 184-188, 194-196, 198, 210-220, 232, 234, 235, 237-242, 248-251, 256-271, 287, 288, 290-293 and 298-337, or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides,
10 recombinant expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

15 In further aspects of the subject invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise: (a) contacting a biological sample with at least one of the above polypeptides; and (b) detecting in the sample the presence of antibodies that bind to the polypeptide or polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.
20 Suitable biological samples include whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine. The diagnostic kits comprise one or more of the above polypeptides in combination with a detection reagent.

The present invention also provides methods for detecting *M. tuberculosis* infection comprising: (a) obtaining a biological sample from a patient;
25 (b) contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the

In a further aspect, the present invention provides a method for detecting *M. tuberculosis* infection in a patient comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with an oligonucleotide probe specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a DNA sequence.

In yet another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figures 2A-D illustrate the reactivity of antisera raised against secretory *M. tuberculosis* proteins, the known *M. tuberculosis* antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with *M. tuberculosis* lysate (lane 2), *M. tuberculosis* secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 6).

Figure 3A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen, TbRa11.

Figure 3B illustrates the stimulation of interferon- γ production in a TbH-

Figure 4 illustrates the reactivity of two representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of bacterial lysate.

Figure 5 shows the reactivity of four representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of the 38 kD antigen.

Figure 6 shows the reactivity of recombinant 38 kD and TbRa11 antigens with sera from *M. tuberculosis* patients, PPD positive donors and normal donors.

Figure 7 shows the reactivity of the antigen TbRa2A with 38 kD negative sera.

Figure 8 shows the reactivity of the antigen of SEQ ID NO: 60 with sera from *M. tuberculosis* patients and normal donors.

Figure 9 illustrates the reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients, PPD positive donors and normal donors as determined by indirect ELISA.

Figure 10 illustrates the reactivity of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors, and with a pool of sera from *M. tuberculosis* patients, as determined both by direct and indirect ELISA.

Figure 11 illustrates the reactivity of increasing concentrations of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors as determined by ELISA.

Figures 12A-E illustrate the reactivity of the recombinant antigens MO-1, MO-2, MO-4, MO-28 and MO-29, respectively, with sera from *M. tuberculosis* patients and from normal donors as determined by ELISA.

SEQ ID NO: 1 is the DNA sequence of TbRa1

- SEQ. ID NO. 4 is the DNA sequence of TbRa12.
SEQ. ID NO. 5 is the DNA sequence of TbRa13.
SEQ. ID NO. 6 is the DNA sequence of TbRa16.
SEQ. ID NO. 7 is the DNA sequence of TbRa17.
5 SEQ. ID NO. 8 is the DNA sequence of TbRa18.
SEQ. ID NO. 9 is the DNA sequence of TbRa19.
SEQ. ID NO. 10 is the DNA sequence of TbRa24.
SEQ. ID NO. 11 is the DNA sequence of TbRa26.
SEQ. ID NO. 12 is the DNA sequence of TbRa28.
10 SEQ. ID NO. 13 is the DNA sequence of TbRa29.
SEQ. ID NO. 14 is the DNA sequence of TbRa2A.
SEQ. ID NO. 15 is the DNA sequence of TbRa3.
SEQ. ID NO. 16 is the DNA sequence of TbRa32.
SEQ. ID NO. 17 is the DNA sequence of TbRa35.
15 SEQ. ID NO. 18 is the DNA sequence of TbRa36.
SEQ. ID NO. 19 is the DNA sequence of TbRa4.
SEQ. ID NO. 20 is the DNA sequence of TbRa9.
SEQ. ID NO. 21 is the DNA sequence of TbRaB.
SEQ. ID NO. 22 is the DNA sequence of TbRaC.
20 SEQ. ID NO. 23 is the DNA sequence of TbRaD.
SEQ. ID NO. 24 is the DNA sequence of YYWCPCG.
SEQ. ID NO. 25 is the DNA sequence of AAMK.
SEQ. ID NO. 26 is the DNA sequence of Tbl-23.
SEQ. ID NO. 27 is the DNA sequence of Tbl-24.
25 SEQ. ID NO. 28 is the DNA sequence of Tbl-25.
SEQ. ID NO. 29 is the DNA sequence of Tbl-28.
SEQ. ID NO. 30 is the DNA sequence of Tbl-29.
SEQ. ID NO. 31 is the DNA sequence of Tbl-3.

SEQ. ID NO. 34 is the DNA sequence of TbM-1.

SEQ. ID NO. 35 is the DNA sequence of TbM-3.

SEQ. ID NO. 36 is the DNA sequence of TbM-6.

SEQ. ID NO. 37 is the DNA sequence of TbM-7.

5 SEQ. ID NO. 38 is the DNA sequence of TbM-9.

SEQ. ID NO. 39 is the DNA sequence of TbM-12.

SEQ. ID NO. 40 is the DNA sequence of TbM-13.

SEQ. ID NO. 41 is the DNA sequence of TbM-14.

SEQ. ID NO. 42 is the DNA sequence of TbM-15.

10 SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

15 SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

SEQ. ID NO. 52 is the DNA sequence of DPEP.

20 SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

25 SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.

SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.

SEQ. ID NO. 61 is the protein sequence of APKT N-terminal Antigen.

SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa1.

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa11.

SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa12.

5 SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa19.

10 SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa24.

SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa26.

SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa28.

SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa29.

SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa2A.

15 SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa3.

SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa32.

SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa35.

SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa36.

SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa4.

20 SEQ. ID NO. 83 is the deduced amino acid sequence of TbRa9.

SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaB.

SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaC.

SEQ. ID NO. 86 is the deduced amino acid sequence of TbRaD.

SEQ. ID NO. 87 is the deduced amino acid sequence of YYWCPG.

25 SEQ. ID NO. 88 is the deduced amino acid sequence of TbAAMK.

SEQ. ID NO. 89 is the deduced amino acid sequence of Tb38-1.

SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-4.

SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-5.

SEQ. ID NO. 94 is the DNA sequence of DPAS.

SEQ. ID NO. 95 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 96 is the DNA sequence of DPV.

SEQ. ID NO. 97 is the deduced amino acid sequence of DPV.

5 SEQ. ID NO. 98 is the DNA sequence of ESAT-6.

SEQ. ID NO. 99 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 100 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 101 is the DNA sequence of TbH-9FL.

SEQ. ID NO. 102 is the deduced amino acid sequence of TbH-9FL

10 SEQ. ID NO. 103 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 104 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 105 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 106 is the deduced amino acid sequence of TbH-9-4.

SEQ. ID NO. 107 is the DNA sequence of Tb38-1F2 IN.

15 SEQ. ID NO. 108 is the DNA sequence of Tb38-1F2 RP.

SEQ. ID NO. 109 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 110 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 111 is the DNA sequence of Tb38-1F3.

SEQ. ID NO. 112 is the deduced amino acid sequence of Tb38-1F3.

20 SEQ. ID NO. 113 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 114 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 115 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 116 is the deduced N-terminal amino acid sequence of AVGS.

SEQ. ID NO. 117 is the deduced N-terminal amino acid sequence of AAMK.

25 SEQ. ID NO. 118 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 119 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of AAES.

SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of DPFP.

SEQ. ID NO. 124 is the protein sequence of DPPD N-terminal Antigen.

SEQ ID NO. 125-128 are the protein sequences of four DPPD cyanogen bromide fragments.

SEQ ID NO. 129 is the N-terminal protein sequence of XDS antigen.

5 SEQ ID NO. 130 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 131 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 132 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 133 is the DNA sequence of TbH-29.

SEQ ID NO. 134 is the DNA sequence of TbH-30.

10 SEQ ID NO. 135 is the DNA sequence of TbH-32.

SEQ ID NO. 136 is the DNA sequence of TbH-33.

SEQ ID NO. 137 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 138 is the predicted amino acid sequence of TbH-30.

SEQ ID NO. 139 is the predicted amino acid sequence of TbH-32.

15 SEQ ID NO. 140 is the predicted amino acid sequence of TbH-33.

SEQ ID NO. 141-146 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO. 147 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

20 SEQ ID NO. 148 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO. 149 is the DNA sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO. 150 is the amino acid sequence of the M. tuberculosis antigen 38 kD.

25 SEQ ID NO. 151 is the DNA sequence of XP14.

SEQ ID NO. 152 is the DNA sequence of XP24.

SEQ ID NO. 153 is the DNA sequence of XP31.

SEQ ID NO. 154 is the 5' DNA sequence of XP33.

SEQ ID NO: 157 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 158 is the DNA sequence of XP27.

SEQ ID NO: 159 is the DNA sequence of XP36.

5 SEQ ID NO: 160 is the 5' DNA sequence of XP4.

SEQ ID NO: 161 is the 5' DNA sequence of XP5.

SEQ ID NO: 162 is the 5' DNA sequence of XP17.

SEQ ID NO: 163 is the 5' DNA sequence of XP30.

SEQ ID NO: 164 is the 5' DNA sequence of XP2.

10 SEQ ID NO: 165 is the 3' DNA sequence of XP2.

SEQ ID NO: 166 is the 5' DNA sequence of XP3.

SEQ ID NO: 167 is the 3' DNA sequence of XP3.

SEQ ID NO: 168 is the 5' DNA sequence of XP6.

SEQ ID NO: 169 is the 3' DNA sequence of XP6.

15 SEQ ID NO: 170 is the 5' DNA sequence of XP18.

SEQ ID NO: 171 is the 3' DNA sequence of XP18.

SEQ ID NO: 172 is the 5' DNA sequence of XP19.

SEQ ID NO: 173 is the 3' DNA sequence of XP19.

SEQ ID NO: 174 is the 5' DNA sequence of XP22.

20 SEQ ID NO: 175 is the 3' DNA sequence of XP22.

SEQ ID NO: 176 is the 5' DNA sequence of XP25.

SEQ ID NO: 177 is the 3' DNA sequence of XP25.

SEQ ID NO: 178 is the full-length DNA sequence of TbH4-XP1.

SEQ ID NO: 179 is the predicted amino acid sequence of TbH4-XP1.

25 SEQ ID NO: 180 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1.

SEQ ID NO: 181 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 182 is a second predicted amino acid sequence encoded by XP36.

SEQ ID NO: 184 is the DNA sequence of RDIF2.

SEQ ID NO: 185 is the DNA sequence of RDIF5.

SEQ ID NO: 186 is the DNA sequence of RDIF8.

SEQ ID NO: 187 is the DNA sequence of RDIF10.

5 SEQ ID NO: 188 is the DNA sequence of RDIF11.

SEQ ID NO: 189 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 190 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 191 is the predicted amino acid sequence of RDIF8.

SEQ ID NO: 192 is the predicted amino acid sequence of RDIF10.

10 SEQ ID NO: 193 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 194 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 195 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 196 is the DNA sequence of RDIF7.

SEQ ID NO: 197 is the predicted amino acid sequence of RDIF7.

15 SEQ ID NO: 198 is the DNA sequence of DIF2-1.

SEQ ID NO: 199 is the predicted amino acid sequence of DIF2-1.

SEQ ID NO: 200-207 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as TbF-2).

20 SEQ ID NO: 208 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 209 is the amino acid sequence of the fusion protein TbF-2.

SEQ ID NO: 210 is the 5' DNA sequence of MO-1.

SEQ ID NO: 211 is the 5' DNA sequence for MO-2.

SEQ ID NO: 212 is the 5' DNA sequence for MO-4.

25 SEQ ID NO: 213 is the 5' DNA sequence for MO-8.

SEQ ID NO: 214 is the 5' DNA sequence for MO-9.

SEQ ID NO: 215 is the 5' DNA sequence for MO-26.

SEQ ID NO: 216 is the 5' DNA sequence for MO-28.

SEQ ID NO: 219 is the 5' DNA sequence for MO-34.

SEQ ID NO: 220 is the 5' DNA sequence for MO-35.

SEQ ID NO: 221 is the predicted amino acid sequence for MO-1.

SEQ ID NO: 222 is the predicted amino acid sequence for MO-2.

5 SEQ ID NO: 223 is the predicted amino acid sequence for MO-4.

SEQ ID NO: 224 is the predicted amino acid sequence for MO-8.

SEQ ID NO: 225 is the predicted amino acid sequence for MO-9.

SEQ ID NO: 226 is the predicted amino acid sequence for MO-26.

SEQ ID NO: 227 is the predicted amino acid sequence for MO-28.

10 SEQ ID NO: 228 is the predicted amino acid sequence for MO-29.

SEQ ID NO: 229 is the predicted amino acid sequence for MO-30.

SEQ ID NO: 230 is the predicted amino acid sequence for MO-34.

SEQ ID NO: 231 is the predicted amino acid sequence for MO-35.

SEQ ID NO: 232 is the determined DNA sequence for MO-10.

15 SEQ ID NO: 233 is the predicted amino acid sequence for MO-10.

SEQ ID NO: 234 is the 3' DNA sequence for MO-27.

SEQ ID NO: 235 is the full-length DNA sequence for DPPD.

SEQ ID NO: 236 is the predicted full-length amino acid sequence for DPPD.

SEQ ID NO: 237 is the determined 5' cDNA sequence for LSER-10.

20 SEQ ID NO: 238 is the determined 5' cDNA sequence for LSER-11.

SEQ ID NO: 239 is the determined 5' cDNA sequence for LSER-12.

SEQ ID NO: 240 is the determined 5' cDNA sequence for LSER-13.

SEQ ID NO: 241 is the determined 5' cDNA sequence for LSER-16.

SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-25.

25 SEQ ID NO: 243 is the predicted amino acid sequence for LSER-10.

SEQ ID NO: 244 is the predicted amino acid sequence for LSER-12.

SEQ ID NO: 245 is the predicted amino acid sequence for LSER-13.

SEQ ID NO: 246 is the predicted amino acid sequence for LSER-16.

SEQ ID NO: 249 is the determined cDNA sequence for LSER-23
SEQ ID NO: 250 is the determined cDNA sequence for LSER-24
SEQ ID NO: 251 is the determined cDNA sequence for LSER-27
SEQ ID NO: 252 is the predicted amino acid sequence for LSER-18
5 SEQ ID NO: 253 is the predicted amino acid sequence for LSER-23
SEQ ID NO: 254 is the predicted amino acid sequence for LSER-24
SEQ ID NO: 255 is the predicted amino acid sequence for LSER-27
SEQ ID NO: 256 is the determined 5' cDNA sequence for LSER-1
SEQ ID NO: 257 is the determined 5' cDNA sequence for LSER-3
10 SEQ ID NO: 258 is the determined 5' cDNA sequence for LSER-4
SEQ ID NO: 259 is the determined 5' cDNA sequence for LSER-5
SEQ ID NO: 260 is the determined 5' cDNA sequence for LSER-6
SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-8
SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-14
15 SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-15
SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-17
SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-19
SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-20
SEQ ID NO: 267 is the determined 5' cDNA sequence for LSER-22
20 SEQ ID NO: 268 is the determined 5' cDNA sequence for LSER-26
SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-28
SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-29
SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-30
SEQ ID NO: 272 is the predicted amino acid sequence for LSER-1
25 SEQ ID NO: 273 is the predicted amino acid sequence for LSER-3
SEQ ID NO: 274 is the predicted amino acid sequence for LSER-5
SEQ ID NO: 275 is the predicted amino acid sequence for LSER-6
SEQ ID NO: 276 is the predicted amino acid sequence for LSER-8

SEQ ID NO: 279 is the predicted amino acid sequence for LSER-17

SEQ ID NO: 280 is the predicted amino acid sequence for LSER-19

SEQ ID NO: 281 is the predicted amino acid sequence for LSER-20

SEQ ID NO: 282 is the predicted amino acid sequence for LSER-22

5 SEQ ID NO: 283 is the predicted amino acid sequence for LSER-26

SEQ ID NO: 284 is the predicted amino acid sequence for LSER-28

SEQ ID NO: 285 is the predicted amino acid sequence for LSER-29

SEQ ID NO: 286 is the predicted amino acid sequence for LSER-30

SEQ ID NO: 287 is the determined cDNA sequence for LSER-9

10 SEQ ID NO: 288 is the determined cDNA sequence for the reverse complement of LSER-6

SEQ ID NO: 289 is the predicted amino acid sequence for the reverse complement of LSER-6

SEQ ID NO: 290 is the determined 5' cDNA sequence for MO-12

15 SEQ ID NO: 291 is the determined 5' cDNA sequence for MO-13

SEQ ID NO: 292 is the determined 5' cDNA sequence for MO-19

SEQ ID NO: 293 is the determined 5' cDNA sequence for MO-39

SEQ ID NO: 294 is the predicted amino acid sequence for MO-12

SEQ ID NO: 295 is the predicted amino acid sequence for MO-12

20 SEQ ID NO: 296 is the predicted amino acid sequence for MO-19

SEQ ID NO: 297 is the predicted amino acid sequence for MO-39

SEQ ID NO: 298 is the determined 5' cDNA sequence for Erdsn-1

SEQ ID NO: 299 is the determined 5' cDNA sequence for Erdsn-2

SEQ ID NO: 300 is the determined 5' cDNA sequence for Erdsn-4

25 SEQ ID NO: 301 is the determined 5' cDNA sequence for Erdsn-5

SEQ ID NO: 302 is the determined 5' cDNA sequence for Erdsn-6

SEQ ID NO: 303 is the determined 5' cDNA sequence for Erdsn-7

SEQ ID NO: 304 is the determined 5' cDNA sequence for Erdsn-8

SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-12
SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-13
SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-14
SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-15
5 SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-16
SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-17
SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-18
SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-21
SEQ ID NO: 315 is the determined 5' cDNA sequence for Erdsn-22
10 SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-23
SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-25
SEQ ID NO: 318 is the determined 3' cDNA sequence for Erdsn-1
SEQ ID NO: 319 is the determined 3' cDNA sequence for Erdsn-2
SEQ ID NO: 320 is the determined 3' cDNA sequence for Erdsn-4
15 SEQ ID NO: 321 is the determined 3' cDNA sequence for Erdsn-5
SEQ ID NO: 322 is the determined 3' cDNA sequence for Erdsn-7
SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-8
SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-9
SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-10
20 SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-12
SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-13
SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-14
SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-15
SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-16
25 SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-17
SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-18
SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-21
SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-22

SEQ ID NO: 337 is the determined cDNA sequence for Erdsn-24

SEQ ID NO: 338 is the determined amino acid sequence for a *M. tuberculosis*
85b precursor homolog

SEQ ID NO: 339 is the determined amino acid sequence for spot 1

5 SEQ ID NO: 340 is a determined amino acid sequence for spot 2

SEQ ID NO: 341 is a determined amino acid sequence for spot 2

SEQ ID NO: 342 is the determined amino acid seq for spot 4

SEQ ID NO: 343 is the sequence of primer PDM-157

SEQ ID NO: 344 is the sequence of primer PDM-160

10 SEQ ID NO: 345 is the DNA sequence of the fusion protein TbF-6

SEQ ID NO: 346 is the amino acid sequence of fusion protein TbF-6

SEQ ID NO: 347 is the sequence of primer PDM-176

SEQ ID NO: 348 is the sequence of primer PDM-175

SEQ ID NO: 349 is the DNA sequence of the fusion protein TbF-8

15 SEQ ID NO: 350 is the amino acid sequence of the fusion protein TbF-8

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to
20 compositions and methods for diagnosing tuberculosis. The compositions of the subject
invention include polypeptides that comprise at least one antigenic portion of a
M. tuberculosis antigen, or a variant of such an antigen that differs only in conservative
substitutions and/or modifications. Polypeptides within the scope of the present
invention include, but are not limited to, soluble *M. tuberculosis* antigens. A "soluble
25 *M. tuberculosis* antigen" is a protein of *M. tuberculosis* origin that is present in
M. tuberculosis culture filtrate. As used herein, the term "polypeptide" encompasses
amino acid chains of any length, including full length proteins (i.e., antigens), wherein
the amino acid residues are linked by covalent bonds (e.g., peptide bonds).

be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be antigenic.

An "antigenic portion" of an antigen (which may or may not be soluble) is a portion that is capable of reacting with sera obtained from an *M. tuberculosis*-infected individual (i.e., generates an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals, in a representative ELISA assay described herein). An "*M. tuberculosis*-infected individual" is a human who has been infected with *M. tuberculosis* (e.g., has an intradermal skin test response to PPD that is at least 0.5 cm in diameter). Infected individuals may display symptoms of tuberculosis or may be free of disease symptoms. Polypeptides comprising at least an antigenic portion of one or more *M. tuberculosis* antigens as described herein may generally be used, alone or in combination, to detect tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

As used herein, a "conservative substitution" is one in which an amino

hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

5 Variants may also, or alternatively, contain other modifications, including the deletion or addition of amino acids that have minimal influence on the antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other 10 sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited 15 nucleotide sequence in having one or more nucleotide deletions, substitutions or additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (*DNA*, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide 20 sequences preferably exhibit at least about 70%, more preferably at least about 80% and most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recite nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 0.5X SSC, 0.2% SDS, hybridizing at 65 °C, 0.5X SSC, 0.2% SDS overnight, 25 followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

In a related aspect, combination, or fusion, polypeptides are disclosed. A 30 fusion polypeptide is a polypeptide in which two or more polypeptides are joined together.

joined directly (i.e., with no intervening amino acids) or may be joined by way of a linker sequence (e.g., Gly-Cys-Gly) that does not significantly diminish the antigenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such
5 antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens may then be evaluated for a desired property, such as the ability to react with sera obtained from an *M. tuberculosis*-infected individual.
10 Such screens may be performed using the representative methods described herein. Antigens may then be partially sequenced using, for example, traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and
15 expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera
20 obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by
25 screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as

therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

5 Regardless of the method of preparation, the antigens described herein are "antigenic." More specifically, the antigens have the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Reactivity may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from infected individuals that is at least three standard
10 deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

 Antigenic portions of *M. tuberculosis* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited
15 therein. Such techniques include screening polypeptide portions of the native antigen for antigenic properties. The representative ELISAs described herein may generally be employed in these screens. An antigenic portion of a polypeptide is a portion that, within such representative assays, generates a signal in such assays that is substantially similar to that generated by the full length antigen. In other words, an antigenic portion
20 of a *M. tuberculosis* antigen generates at least about 20%, and preferably about 100%, of the signal induced by the full length antigen in a model ELISA as described herein.

 Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated
25 using techniques well known in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963).

according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

5 Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available
10 filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

 Any of a variety of expression vectors known to those of ordinary skill in
15 the art may be employed to express recombinant polypeptides as described herein. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian
20 cell line, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

 In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides
25 are at least about 30% pure, more preferably at least about 90% pure and most preferably at least about 99% pure. For use in the methods described herein, however, such substantially pure polypeptides may be combined

 For copyright purposes, this document is hereby published in perpetuity.

antigen (or a variant of such an antigen), where the antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- 5 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- 10 (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);
- 15 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);
- 20 (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser (SEQ ID NO: 124);
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp (SEQ ID NO: 130) or
- 25 (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly (SEQ ID NO: 131)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence

encoding the antigen identified as (a) above is provided in SEQ ID NO: 96; its deduced amino acid sequence is provided in SEQ ID NO: 97. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID NO: 24, a DNA sequence corresponding to antigen (c) is provided in SEQ ID NO: 25 and a DNA sequence corresponding to
5 antigen (f) is disclosed in SEQ ID NO: 94 and its deduced amino acid sequence is provided in SEQ ID NO: 95.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only
10 in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

15 wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen of (n) above is provided in SEQ ID NO: 235, with the corresponding predicted full-length amino acid sequence being provided in SEQ ID NO: 236.

20 In other specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID NOS 1, 2, 4, 10, 13-25, 52, 94 and 96, (b) the complements of such DNA sequences, or (c) DNA sequences
25 substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) which may or may not be soluble that comprises one or

242, 248-251, 256-271, 287, 288, 290-293 and 298-337, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In a related aspect, the present invention provides fusion proteins
5 comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described in Andersen and Hansen, *Infect Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID NOS: 98 and 99), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a
10 linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or
15 without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the
20 second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and
25 (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids such as Thr and Ala are also acceptable.

Natl. Acad. Sci. USA 83:8258-8562, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric hindrance.

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose tuberculosis. In this aspect, methods are provided for detecting *M. tuberculosis* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, may be included. As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply. The polypeptide(s) are used in an assay, as described below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut-off value. The presence of such antibodies indicates previous sensitization to mycobacterial antigens which may be indicative of tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (i.e., one component polypeptide will tend to detect infection in samples where the infection would not be detected by another component polypeptide). Complementary polypeptides may generally be identified by using each polypeptide individually to evaluate serum samples obtained from a series of patients known to be infected with *M. tuberculosis*. After determining which samples test positive (as described below) with each polypeptide, combinations of two or more polypeptides may be formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately

polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. See, e.g.,
5 Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include
10 antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (e.g., in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the
15 binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane.
20 Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of
25 techniques known to those of ordinary skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "bound" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the

be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging
5 from about 10 ng to about 1 μ g, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the
10 polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (see, e.g., Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent
15 assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized
20 antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically
25 blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The immobilized polypeptide is then incubated with the sample, and antibody is allowed to bind to the antigen. The sample may be diluted with a suitable

detect the presence of antibody within a *M. tuberculosis*-infected sample. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily
5 determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any
10 compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors,
15 inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups, biotin and colloidal particles, such as colloidal gold and selenium. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (e.g., Zymed Laboratories,
20 San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time.
25 Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic

radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of anti-*M. tuberculosis* antibodies
5 in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above
10 the predetermined cut-off value is considered positive for tuberculosis. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive
15 rates (i.e., sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (i.e., the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the
20 cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through
25 or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (e.g., protein A colloidal gold) then binds to the antibody-polypeptide complex.

strip test format, one end of the membrane to which polypeptide is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing detection reagent and to the area of immobilized polypeptide. Concentration of detection reagent at the polypeptide indicates the presence of anti-
5 *M. tuberculosis* antibodies in the sample. Typically, the concentration of detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be
10 sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 µg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (e.g., one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use
15 with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

In yet another aspect, the present invention provides antibodies to the inventive polypeptides. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A
20 Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the antigenic polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep and goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response
25 may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies

Monoclonal antibodies specific for the antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process, for example, an affinity chromatography step.

Antibodies may be used in diagnostic tests to detect the presence of *M. tuberculosis* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *M. tuberculosis* infection in a patient.

thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify *M. tuberculosis*-specific cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for a DNA molecule encoding a polypeptide of the present invention.

5 The presence of the amplified cDNA is then detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes specific for a DNA molecule encoding a polypeptide of the present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

10 As used herein, the term "oligonucleotide primer/probe specific for a DNA molecule" means an oligonucleotide sequence that has at least about 80%, preferably at least about 90% and more preferably at least about 95%, identity to the DNA molecule in question. Oligonucleotide primers and/or probes which may be usefully employed in the inventive diagnostic methods preferably have at least about
15 10-40 nucleotides. In a preferred embodiment, the oligonucleotide primers comprise at least about 10 contiguous nucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Preferably, oligonucleotide probes for use in the inventive diagnostic methods comprise at least about 15 contiguous oligonucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Techniques for
20 both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis *et al.* *Ibid*; Ehrlich, *Ibid*). Primers or probes may thus be used to detect *M. tuberculosis*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES

FROM *M. TUBERCULOSIS* CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was then filtered through a 0.2 μ filter into a sterile 4 L bottle. NaN_3 was then added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was then dialyzed into 0.1% ammonium bicarbonate using a 5,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were then dialyzed against 0.01 mM 1,3-bis[tris(hydroxymethyl)methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for ion exchange chromatography. Fractionation was performed using

Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were
5. dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate
10 was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD
15 skin test positive and whose T cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 µg/ml gentamicin. Purified polypeptides were added in duplicate at concentrations of 0.5 to 10 µg/mL. After six days of culture in 96-well round-bottom plates in a volume of 200 µl, 50 µl of medium
20 was removed from each well for determination of IFN-γ levels, as described below. The plates were then pulsed with 1 µCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

25 IFN-γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN-γ (Chemicon) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 8% (W/V) non fat dried milk for 1 hour at room

room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN- γ serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Jackson Labs.) was added at a 1:2000
5 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone,
10 plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto Biobrene™ (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced
15 from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- 20 (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Ihr-Naa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 54);
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 55);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Ihr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 56);
25
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 57);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Naa-Ala-Val

- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 59);
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala (SEQ ID NO: 60); and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 61);

wherein Xaa may be any amino acid.

An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20 μ l of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer/Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 μ l/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to have a molecular weight of 12,054 Kd (by mass spectrometry) and the following N-terminal sequence.

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID NO: 62)

This polypeptide was shown to induce proliferation and IFN- γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following analysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using ion exchange chromatography on a Poros Q1 column 4.6 x 100 mm.

were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended in 80 µl of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131), wherein Xaa may be any amino acid

Using the assays described above, these polypeptides were shown to induce proliferation and IFN- γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and (g) above were obtained by screening a *M. tuberculosis* genomic library using ³²P end

corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID NO: 96. The polypeptide encoded by SEQ ID NO: 96 is provided in SEQ ID NO: 97. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID NO: 52. The polypeptide encoded by SEQ ID NO: 52 is provided in SEQ ID NO: 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID NO: 24, and the screen performed with a probe corresponding to antigen (c) identified a clone having the sequence provided in SEQ ID NO: 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen an *M. tuberculosis* library and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID NO: 94).

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1.

TABLE 1
RESULTS OF PBMC PROLIFERATION AND IFN- γ ASSAYS

Sequence	Proliferation	IFN- γ
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, as SI of 4-8 or 2-4 at a concentration of 1 μ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (i) was found to have a high SI (++++) for one donor and lower SI (--- and -) for the two other donors in both proliferation and IFN- γ assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon- γ production.

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Desiccated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with

DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic *M. tuberculosis* Erdman strain library using labeled degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO: 198. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 199. Comparison of these sequences with those in the genbank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING *M. tuberculosis* ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* SUPERNATANT

Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the
5 Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 100 µg muramyl dipeptide
10 (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg protein antigen. The anti-sera were used to screen the expression library as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor
15 Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in *M. tuberculosis*. Proteins were induced by IPTG
20 and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 187:1527-1537, 1995. Representative partial sequences of DNA molecules identified in this screen are provided in SEQ ID NOS 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID NOS 64-88.

On comparison of these sequences with known sequences in the gene
25 bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID NOS 77, 69, 71, 70) show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA2A (SEQ ID NOS 77) is identical to

previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRA19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID NOS: 64, 78, 82, 83, 65, 68, 76, 72, 76, 79, 81, 80, 67, respectively). The clone
5 TbRa24 is overlapping with clone TbRa29

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial Sau3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with
10 active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (i.e., TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active
15 pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified.
20 Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty-two clones were purified. Of these, 11 represented sequences that had not been previously identified in human *M. tuberculosis*. (Table 1)

NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID NOS. 89-93. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 107, 108, 111, 113, and 114). (SEQ. ID NOS. 107 and 108 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames were deduced in Tb38-1F2: one corresponds to Tb37FL (SEQ. ID. NO. 109), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 110). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 112. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 101), which may be the homologue of TbH-9 (H37Ra), TbH-9-1 (SEQ. ID NO. 103), and TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8. The deduced amino acid sequences for these three clones are presented in SEQ. ID NOS. 102, 104 and 106.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined

TbH-33) are provided in SEQ ID NO: 133-136, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 137-140, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E. coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ¹²⁵I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 2.

TABLE 2

Antigen	Human <i>M. tb</i> Sera	Anti-lacZ Sera
TbH-29	45 Kd	45 Kd
TbH-30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd	16 Kd

the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 2A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 µg of *M. tuberculosis* lysate; 3) 5 µg secretory proteins; 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger than the native protein. In Figure 2D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones with *M. tuberculosis* lysate.

The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 3A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 3B shows the production of IFN- γ by a second TbH-9-specific T cell clone (designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG - A - M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID NOS. 151-153, respectively, with the 5' and 3' DNA sequences for XP22 being provided in SEQ ID NOS. 154 and 155, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID NO: 156. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID NO: 157.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those

NOS: 158 and 159, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID NOS: 160-163, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID NOS: 164 and 165; 166 and 167; 168 and 169; 170 and 171; 172 and 173; 174 and 175; and 176 and 177, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID NO: 178. This DNA sequence was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 179. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 180. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID NOS: 181 and 182, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 183.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. Recombinant XP1 was found to stimulate cell proliferation and IFN- γ production in T cells isolated from an *M. tuberculosis*-immune donors.

D. USE OF A LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* H37Rv strain, randomly sheared and used to construct an expression library employing the Lambda screen expression system (Novagen, Madison, WI), as described below in Example 6. Pooled serum obtained from *M. tuberculosis*-infected patients and that was shown to react with *M. tuberculosis* lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being anti-mouse IgG.

(hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined 5' cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID NO: 237-242, respectively, with the corresponding predicted amino acid sequences for LSER-10, LSER-12, LSER-13, LSER-16 and LSER-25 being provided in SEQ ID NO: 243-247, respectively. The determined full-length cDNA sequences for LSER-18, LSER-23, LSER-24 and LSER-27 are shown in SEQ ID NO: 248-251, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 252-255. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 256-271, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 272-286, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 287. The reverse complement of LSER-9 (SEQ ID NO: 288) was found to encode the predicted amino acid sequence of SEQ ID NO: 289.

EXAMPLE 2 PREPARATION OF M. TUBERCULOSIS SOLUBLE ANTIGENS USING RABBIT ANTISERA RAISED AGAINST M. TUBERCULOSIS FRACTIONATED PROTEINS

M. tuberculosis lysate was prepared as described above in Example 1. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera were prepared against the antigenic fractions.

Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, six (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID NOS: 184-188, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NOS: 189-193, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID NOS: 194 and 195, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID NOS: 196 and 197, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. These antigens were found to stimulate cell proliferation and IFN- γ production in T cells isolated from *M. tuberculosis*-immune donors.

EXAMPLE 4

PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows:

PPD was prepared as published with some modification (Seibert, E. et

Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100°C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 μ l/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N terminal sequence shown in SEQ ID NO: 124. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologues. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID

sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 235, with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 236.

5

EXAMPLE 5

USE OF SERA FROM TUBERCULOSIS-INFECTED MONKEYS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

10

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 210-220, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 221-231. The full-length DNA sequence of the clone MO-10 is provided in SEQ ID NO: 232, with the corresponding predicted amino acid sequence being provided in SEQ ID NO: 233. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 234.

Clones MO-1, MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTC123. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All clones

and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified 14 kDa *M. tuberculosis* heat shock protein

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinafter referred to as MO-12, MO-13, MO-19 and MO-39. The determined 5' cDNA sequences for these clones are provided in SEQ ID NO: 290-293, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 294-297, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

EXAMPLE D

ISOLATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS BY SCREENING OF A NOVEL EXPRESSION LIBRARY

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRa11, TbRa3, Tb38-1, TbH4, TbF and 38 kD

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase (Boehringer

extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the λ Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1, Erdsn2, Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 298-317, respectively, with the determined 3' cDNA sequences for Erdsn1, Erdsn2, Erdsn4, Erdsn5, Erdsn7-Erdsn10, Erdsn12-Erdsn18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 318-336, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 337. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 304, 311, 313-315, 317, 319, 324, 326, 329, 331, 333, 335 and 337. The sequences of SEQ ID NO: 298-303, 305-310, 312, 316, 318, 320-321, 324-326, 328, 330, 332, 334 and 336 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

EXAMPLE 7

ISOLATION OF SOLUBLE *M. TUBERCULOSIS* ANTIGENS

USING MASS SPECTROMETRY

This example illustrates the use of mass spectrometry to identify soluble *M. tuberculosis* antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined

the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity, although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jnl. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1, 2 and 4) are provided in SEQ ID NO: 339, 340-341 and 342, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified *PcK-1*, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAs, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein *Kat G*. To the best of the inventors' knowledge, neither *PcK-1* nor the two DNAs have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

EXAMPLE 5

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide

attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-*t*-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

This procedure was used to synthesize a TbM-1 peptide that contains one and a half repeats of a TbM-1 sequence. The TbM-1 peptide has the sequence GCGDRSGGNLDQIRLRDRSGGNL (SEQ ID NO: 63).

EXAMPLE 9

USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

This Example illustrates the diagnostic properties of several representative antigens.

Assays were performed in 96-well plates were coated with 200 ng antigen diluted to 50 μ L in carbonate coating buffer, pH 9.6. The wells were coated overnight at 4°C (or 2 hours at 37°C). The plate contents were then removed and the wells were blocked for 2 hours with 200 μ L of PBS/1% BSA. After the blocking step, the wells were washed five times with PBS/0.1% Tween 20™. 50 μ L sera, diluted 1:100 in PBS/0.1% Tween 20™/0.1% BSA, was then added to each well and incubated for 30 minutes at room temperature. The plates were then washed again five times with PBS/0.1% Tween 20™.

μL of the diluted conjugate was added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells were washed five times with PBS/0.1% Tween 20™. 100 μL of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) was added, undiluted, and
5 incubated for about 15 minutes. The reaction was stopped with the addition of 100 μL of 1 N H₂SO₄ to each well, and the plates were read at 450 nm.

Figure 4 shows the ELISA reactivity of two recombinant antigens isolated using method A in Example 3 (TbRa3 and TbRa9) with sera from *M. tuberculosis* positive and negative patients. The reactivity of these antigens is
10 compared to that of bacterial lysate isolated from *M. tuberculosis* strain H37Ra (Difco, Detroit, MI). In both cases, the recombinant antigens differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 56 out of 87 positive sera, and TbRa9 detected 111 out of 165 positive sera.

Figure 5 illustrates the ELISA reactivity of representative antigens
15 isolated using method B of Example 3. The reactivity of the recombinant antigens TbH4, TbH12, Tb38-1 and the peptide TbM-1 (as described in Example 4) is compared to that of the 38 kD antigen described by Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989. Again, all of the polypeptides tested differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbH4
20 detected 97 out of 126 positive sera, TbH12 detected 80 out of 125 positive sera, 38-1 detected 91 out of 101 positive sera and the TbM-1 peptide detected 25 out of 30 positive sera.

The reactivity of four antigens (TbRa3, TbRa9, TbH4 and TbH12) with sera from a group of *M. tuberculosis* infected patients with differing reactivity in the
25 acid fast stain of sputum (Smithwick and David, *Tubercle* 52:226, 1971) was also examined, and compared to the reactivity of *M. tuberculosis* lysate and the 38 kD antigen. The results are presented in Table 3, below.

TABLE 3
 REACTIVITY OF ANTIGENS WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-2	----	1.853	0.634	0.998	1.022	1.030	1.314
Tb01B93I-19	-----	2.657	2.322	0.608	0.837	1.857	2.335
Tb01B93I-8	---	2.703	0.527	0.492	0.281	0.501	2.002
Tb01B93I-10	---	1.565	1.301	0.685	0.216	0.448	0.458
Tb01B93I-11	---	2.817	0.697	0.509	0.301	0.173	2.608
Tb01B93I-15	---	1.28	0.283	0.808	0.218	1.537	0.811
Tb01B93I-16	---	2.908	>3	0.899	0.441	0.593	1.080
Tb01B93I-25	---	0.395	0.131	0.335	0.211	0.107	0.948
Tb01B93I-87	---	2.653	2.432	2.282	0.977	1.221	0.857
Tb01B93I-89	---	1.912	2.370	2.436	0.876	0.520	0.952
Tb01B94I-108	---	1.639	0.341	0.797	0.368	0.654	1.798
Tb01B94I-201	---	1.721	0.410	0.661	0.137	0.064	0.692
Tb01B93I-88	--	1.930	1.269	2.519	1.381	0.214	0.530
Tb01B93I-92	--	2.155	2.329	2.78	0.685	0.997	2.527
Tb01B94I-109	--	0.993	0.620	0.574	0.441	0.5	2.558
Tb01B94I-210	---	2.777	>3	0.393	0.367	1.004	1.315
Tb01B94I-224	---	2.913	0.476	0.251	1.297	1.990	0.256
Tb01B93I-9	-	2.649	0.278	0.210	0.140	0.181	1.586

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-22	+	0.714	0.451	2.082	0.285	0.269	1.159
Tb01B93I-31	+	0.956	0.490	1.019	0.812	0.176	1.293
Tb01B93I-32	-	2.261	0.786	0.668	0.273	0.535	0.405
Tb01B93I-52		0.658	0.114	0.434	0.330	0.273	1.140
Tb01B93I-99	-	2.118	0.584	1.62	0.119	0.977	0.729
Tb01B94I-130		1.349	0.224	0.86	0.282	0.583	2.146
Tb01B94I-131	-	0.685	0.324	1.173	0.059	0.118	1.431
AT4-0070	Normal	0.072	0.043	0.092	0.071	0.040	0.039
AT4-0105	Normal	0.397	0.121	0.118	0.103	0.078	0.390
3-15/94-1	Normal	0.227	0.064	0.098	0.026	0.001	0.228
4-15/93-2	Normal	0.114	0.240	0.071	0.034	0.041	0.264
5-26/94-4	Normal	0.089	0.259	0.096	0.046	0.008	0.053
5-26/94-3	Normal	0.139	0.093	0.085	0.019	0.067	0.01

Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 23 out of 27 positive sera. TbRa9 detected 22 out of 27. TbH4 detected 18 out of 27 and TbH12 detected 15 out of 27. If used in combination, these four antigens would have a theoretical sensitivity of 27 out of 27, indicating that these antigens should complement each other in the serological detection of *M. tuberculosis* infection. In addition, several of the recombinant antigens detected positive sera that were not detected using the 38 kD antigen, indicating that these antigens may be complementary to the 38 kD antigen.

The results are shown in Figure 6 which indicates that TbRa11, while being negative with sera from PPD positive and normal donors, detected sera that were negative with the 38 kD antigen. Of the thirteen 38 kD negative sera tested, nine were positive with TbRa11, indicating that this antigen may be reacting with a sub-group of 38 kD antigen negative sera. In contrast, in a group of 38 kD positive sera where TbRa11 was reactive, the mean OD 450 for TbRa11 was lower than that for the 38 kD antigen. The data indicate an inverse relationship between the presence of TbRa11 activity and 38 kD positivity.

The antigen TbRa2A was tested in an indirect ELISA using initially 50 μ l of serum at 1:100 dilution for 30 minutes at room temperature followed by washing in PBS Tween and incubating for 30 minutes with biotinylated Protein A (Zymed, San Francisco, CA) at a 1:10,000 dilution. Following washing, 50 μ l of streptavidin-horseradish peroxidase (Zymed) at 1:10,000 dilution was added and the mixture incubated for 30 minutes. After washing, the assay was developed with TMB substrate as described above. The reactivity of TbRa2A with sera from *M. tuberculosis* patients and normal donors is shown in Table 4. The mean value for reactivity of TbRa2A with sera from *M. tuberculosis* patients was 0.444 with a standard deviation of 0.309. The mean for reactivity with sera from normal donors was 0.109 with a standard deviation of 0.029. Testing of 38 kD negative sera (Figure 7) also indicated that the TbRa2A antigen was capable of detecting sera in this category.

TABLE 4
REACTIVITY OF TBRA2A WITH SERA FROM *M. TUBERCULOSIS* PATIENTS AND FROM
NORMAL DONORS

Serum ID	Status	OD 450
Tb85	TB	0.680
Tb86	TB	0.450
Tb87	TB	0.263

Tb93	TB	0.232
Tb94	TB	0.333
Tb95	TB	0.435
Tb96	TB	0.284
Tb97	TB	0.320
Tb99	TB	0.328
Tb100	TB	0.817
Tb101	TB	0.607
Tb102	TB	0.191
Tb103	TB	0.228
Tb107	TB	0.324
Tb109	TB	1.572
Tb112	TB	0.338
DL4-0176	Normal	0.036
AT4-0043	Normal	0.126
AT4-0044	Normal	0.130
AT4-0052	Normal	0.135
AT4-0053	Normal	0.133
AT4-0062	Normal	0.128
AT4-0070	Normal	0.088
AT4-0091	Normal	0.108
AT4-0100	Normal	0.106
AT4-0105	Normal	0.108
AT4-0109	Normal	0.105

The reactivity of the recombinant antigen (g) (SEQ ID NO: 60) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. Figure 8 shows the results of the titration of antigen (g) with four *M. tuberculosis* positive sera that were all reactive with the 38 kD antigen and with four donor sera. All four positive sera were reactive with antigen (g).

The reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients, PPD positive donors and normal donors was determined by indirect ELISA as described above. The results are shown in Figure 9. TbH-29 detected 30 out of 60 *M. tuberculosis* sera, 2 out of 8 PPD positive sera and 2 out of 27 normal sera.

Figure 9 shows the results of the titration of antigen (g) with *M. tuberculosis* patients. The mean

OD 450 was demonstrated to be higher with sera from *M. tuberculosis* patients than from normal donors, with the mean OD 450 being significantly higher in the indirect ELISA than in the direct ELISA. Figure 11 is a titration curve for the reactivity of recombinant TbH-33 with sera from *M. tuberculosis* patients and from normal donors showing an increase in OD 450 with increasing concentration of antigen.

The reactivity of the recombinant antigens RDIF6, RDIF8 and RDIF10 (SEQ ID NOS: 184-187, respectively) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. RDIF6 detected 6 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; RDIF8 detected 14 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; and RDIF10 detected 4 out of 27 *M. tuberculosis* sera and 1 out of 15 normal sera. In addition, RDIF10 was found to detect 0 out of 5 sera from PPD-positive donors.

The antigens MO-1, MO-2, MO-4, MO-28 and MO-29 described above in Example 5, were expressed in *E. coli* and purified using a hexahistidine tag. The reactivity of these antigens with both *M. tuberculosis* positive and negative sera was examined by ELISA as described above. Titration curves showing the reactivity of MO-1, MO-2, MO-4, MO-28 and MO-29 at different solid phase coat levels when tested against four *M. tuberculosis* positive sera and four *M. tuberculosis* negative sera are shown in Figs. 12A-E, respectively. Three of the clones, MO-1, MO-2 and MO-29 were further tested on panels of HIV positive/tuberculosis (HIV TB) positive and extrapulmonary sera. MO-1 detected 3/20 extrapulmonary and 2/38 HIV TB sera. On the same sera groups, MO-2 detected 2/20 and 10/38, and MO-29 detected 2/20 and 8/38 sera. In combination these three clones would have detected 4/20 extrapulmonary sera and 16/38 HIV TB sera. In addition, MO-1 detected 6/17 sera that had previously been shown only to react with *M. tuberculosis* lysate and not with either 38 kD or with other antigens of the subject invention.

EXAMPLE 10PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was
5 prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified
by PCR in order to facilitate their fusion and the subsequent expression of the fusion
protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform
PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 141 and 142), PDM-57 and
10 PDM-58 (SEQ ID NO: 143 and 144), and PDM-69 and PDM-60 (SEQ ID NO: 145-
146), respectively. In each case, the DNA amplification was performed using 10 µl
10X Pfu buffer, 2 µl 10 mM dNTPs, 2 µl each of the PCR primers at 10 µM
concentration, 81.5 µl water, 1.5 µl Pfu DNA polymerase (Stratagene, La Jolla, CA)
and 1 µl DNA at either 70 ng/µl (for TbRa3) or 50 ng/µl (for 38 kD and Tb38-1). For
15 TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C
for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at
96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15
sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C
for 2 min was followed by 10 cycles of 96°C for 15 sec, 68°C for 15 sec and 72°C for
20 1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by
72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned
directly into pT7⁺L2 IL-1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment
was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and
25 then digested with EcoRI for direct cloning into the pT7⁺L2Ra3-1 vector which was
digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and
EcoRI and directly subcloned into pT7⁺L2Ra3 38kD-17 digested with the same
enzymes. The whole fusion was then transferred to pET28b (Novagen, Madison, WI).

The expression construct was transformed to BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD₅₆₀ of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbRa3-38 kD-Tb38-1) are provided in SEQ ID NO: 147 and 148, respectively.

A fusion protein containing the two antigens TbH-9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 151.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO: 145) and PDM-83 (SEQ ID NO: 200) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3 38kD Tb38-1A fusion was then transferred to pET28b using NdeI and EcoRI sites.

Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 203 and 204, respectively.

10 A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 343 and 344, respectively) and 2 µl DNA at 100 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 345 and 346, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

25 38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 347 and 348, respectively), and 1 µl PET28Ra3/38kD/38-1 Ra2A-12 DNA at 110 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 72 °C for 15 sec and 72 °C for 5 min.

then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pT7ΔL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pT7ΔL2 construct was designed to have a 5 MGHHHHHH amino acid coding region in frame just 5' of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. Ligations were done with a ligation kit from Panvera (Madison, WI). The resulting construct was digested with NdeI (New England Biolabs) and Eco RI, and transferred to a modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 349 and 350, respectively.

EXAMPLE 11

USE OF *M. TUBERCULOSIS* FUSION PROTEINS FOR SERODIAGNOSIS OF TUBERCULOSIS

The effectiveness of the fusion protein TbRa3-38 and Tb38-1, prepared as described above, in the serodiagnosis of tuberculosis infection was examined by ELISA.

The ELISA protocol was as described above in Example 6, with the fusion protein being coated at 200 ng/well. A panel of sera was chosen from a group of tuberculosis patients previously shown, either by ELISA or by culture, to be

all three epitopes functioned with the fusion protein. As shown in Table 5, all four sera that reacted with TbRa3 only were detectable with the fusion protein. Three sera that reacted only with Tb38-1 were also detectable, as were two sera that reacted with 38 kD alone. The remaining 15 sera were all positive with the fusion protein based on a cut-off in the assay of mean negatives +3 standard deviations. This data demonstrates the functional activity of all three epitopes in the fusion protein.

TABLE 5
REACTIVITY OF TRI-PEPTIDE FUSION PROTEIN WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Serum ID	Status	ELISA and/or Western Blot Reactivity with Individual proteins			Fusion Recombinant OD 450	Fusion Recombinant Status
		38kd	Tb38-1	TbRa3		
01B93I-40	TB	-	-	-	0.413	-
01B93I-41	TB	-	-	-	0.392	-
01B93I-29	TB	-	-	-	2.217	-
01B93I-109	TB	-	-	-	0.522	-
01B93I-132	TB	-	-	-	0.937	-
5004	TB	-	-	-	1.098	-
15004	TB	-	-	-	2.077	-
39004	TB	-	-	-	1.675	-
68004	TB	-	-	-	2.388	-
99004	TB	-	-	-	0.607	-
107004	TB	-	-	-	0.667	-
92004	TB	-	-	-	1.070	-
97004	TB	-	-	-	1.152	-
118004	TB	-	-	-	2.694	-
173004	TB	-	-	-	3.258	-
175004	TB	-	-	-	2.514	-
274004	TB	-	-	-	-	-

308004	TB	-	-	-	3.338	+
314004	TB	-	+	-	1.362	+
317004	TB	-	-	-	0.763	+
312004	TB	-	-	+	1.079	+
D176	PPD	-	-	-	0.145	-
D162	PPD	-	-	-	0.073	-
D161	PPD	-	-	-	0.097	-
D27	PPD	-	-	-	0.082	-
A6-124	NORMAL	-	-	-	0.053	-
A6-125	NORMAL	-	-	-	0.087	-
A6-126	NORMAL	-	-	-	0.346	-
A6-127	NORMAL	-	-	-	0.064	-
A6-128	NORMAL	-	-	-	0.034	-
A6-129	NORMAL	-	-	-	0.037	-
A6-130	NORMAL	-	-	-	0.057	-
A6-131	NORMAL	-	-	-	0.054	-
A6-132	NORMAL	-	-	-	0.022	-
A6-133	NORMAL	-	-	-	0.147	-
A6-134	NORMAL	-	-	-	0.101	-
A6-135	NORMAL	-	-	-	0.066	-
A6-136	NORMAL	-	-	-	0.054	-
A6-137	NORMAL	-	-	-	0.065	-
A6-138	NORMAL	-	-	-	0.041	-
A6-139	NORMAL	-	-	-	0.103	-
A6-140	NORMAL	-	-	-	0.212	-
A6-141	NORMAL	-	-	-	0.156	-
A6-142	NORMAL	-	-	-	0.051	-

The reactivity of the fusion protein TbF-2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 6) demonstrate that all four antigens function independently in the ELISA assay.

TABLE 6
REACTIVITY OF TbF-2 FUSION PROTEIN WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa3	Tb38-1	DPEP
B931-40	TB	0.57	-	0.321	+	-	-	-	-
B931-41	TB	0.601	+	0.396	-	+	+	+	+
B931-109	TB	0.494	-	0.404	+	+	+	+	+
B931-132	TB	1.502	-	1.292	-	+	+	+	+
5004	TB	1.806	-	1.600	-	+	+	+	+
15004	TB	2.862	-	2.468	-	+	+	+	+
39004	TB	2.443	-	1.722	-	+	+	+	+
08004	TB	2.371	-	2.575	-	+	+	+	+
99004	TB	0.691	-	0.971	-	+	+	+	+
107004	TB	0.875	-	0.732	-	+	+	+	+
92004	TB	1.632	-	1.394	-	+	+	+	+
97004	TB	1.491	-	1.979	+	+	+	+	+
118004	TB	3.182	-	3.045	+	+	+	+	+
173004	TB	3.644	-	3.578	-	+	+	+	+
175004	TB	3.332	-	2.916	-	+	+	+	+
274004	TB	3.696	-	3.716	-	+	+	+	+
276004	TB	3.243	-	2.56	-	+	+	+	+
282004	TB	1.249	-	1.234	-	+	+	+	+
289004	TB	1.373	-	1.17	-	+	+	+	+
308004	TB	3.708	-	3.355	-	+	+	+	+
314004	TB	1.663	-	1.290	-	+	+	+	+
317004	TB	1.163	-	0.92	-	+	+	+	+
312004	TB	1.709	-	1.453	-	+	+	+	+
380004	TB	0.228	-	0.461	-	+	+	+	+
451004	TB	0.18	-	0.2	-	+	+	+	+
475004	TB	0.188	-	0.469	+	+	+	+	+
410004	TB	0.384	-	2.392	-	+	+	+	+
411004	TB	0.306	-	0.874	-	+	+	+	+
421004	TB	0.337	-	1.426	-	+	+	+	+
28004	TB	0.047	-	0.196	-	+	+	+	+
A6-87	Normal	0.094	-	0.093	-	+	+	+	+
A6-88	Normal	0.214	-	0.19	-	+	+	+	+
A6-89	Normal	0.248	-	0.125	-	+	+	+	+
A6-90	Normal	0.179	-	0.206	-	+	+	+	+
A6-91	Normal	0.138	-	0.151	-	+	+	+	+
A6-92	Normal	0.064	-	0.097	-	+	+	+	+
A6-93	Normal	0.072	-	1.098	-	+	+	+	+
A6-94	Normal	0.072	-	0.064	-	+	+	+	+
A6-95	Normal	0.125	-	0.159	-	+	+	+	+
A6-96	Normal	0.072	-	0.064	-	+	+	+	+

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the
5 construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the
10 invention.

CLAIMS

We claim:

1. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or in modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- (b) Ala-Val-Glu-Ser-Gly-Met-Ile-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 17);
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123); and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly (SEQ ID NO: 131)

wherein Xaa may be any amino acid

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

3. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an antigenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 196, 235, 237-242, 248-251, 290-293, 304, 311, 313-315, 317, 319, 323, 324, 328, 330, 332, 334 and 336, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 196, 235, 237-242, 248-251, 290-293, 304, 311, 313-315, 317, 319, 323, 324, 328, 330, 332, 334 and 336, or a complement thereof under moderately stringent conditions.

6. A recombinant expression vector comprising a DNA molecule according to claim 5.
7. A host cell transformed with an expression vector according to claim 6.
8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.
9. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
 - (a) contacting a biological sample with one or more polypeptides according to any of claims 1-4; and
 - (b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.
10. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
 - (a) contacting a biological sample with a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
 - (b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.
11. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
 - (a) contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NO: 129 and 130;

sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and

(h) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

12. The method of any one of claims 9-11 wherein step (a) additionally comprises contacting the biological sample with a 38 kD *M. tuberculosis* antigen and step (b) additionally comprises detecting in the sample the presence of antibodies that bind to the 38 kD *M. tuberculosis* antigen.

13. The method of any one of claims 9-11 wherein the polypeptide(s) are bound to a solid support.

14. The method of claim 13 wherein the solid support comprises nitrocellulose, latex or a plastic material.

15. The method of any one of claims 9-11 wherein the biological sample is selected from the group consisting of whole blood, serum, plasma, saliva, cerebrospinal fluid and urine.

16. The method of claim 15 wherein the biological sample is whole blood or serum.

17. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a

(b) detecting in the sample a DNA sequence that amplifies in the presence of the oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

18. The method of claim 17, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA molecule according to claim 5.

19. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

20. The method of claim 19, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

21. The method of claims 17 or 19 wherein the biological sample selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. A method for detecting *M. tuberculosis* infection in a biological

- (a) contacting the sample with one or more oligonucleotide probes specific for a DNA molecule according to claim 5; and
- (b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

23. The method of claim 22 wherein the probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

24. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

- (a) contacting the sample with one or more oligonucleotide probes specific for a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and
- (b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

25. The method of claim 24 wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

26. The method of claims 22 or 24 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

27. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to any one of claims 1-4; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

28. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

29. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

30. The method of any one of claims 27-29 wherein the binding agent is a monoclonal antibody.

31. The method of any one of claims 27-29 wherein the binding agent is a polyclonal antibody.
32. A diagnostic kit comprising:
- (a) one or more polypeptides according to any of claims 1-4; and
 - (b) a detection reagent.
33. A diagnostic kit comprising:
- (a) one or more polypeptides having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
 - (b) a detection reagent.
34. A diagnostic kit comprising:
- (a) one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, and
 - (b) a detection reagent.
35. The kit of any one of claims 32-34 wherein the polypeptide(s) are immobilized on a solid support.
36. The kit of claim 35 wherein the solid support comprises nitrocellulose, latex, or a plastic material.

38. The kit of claim 37 wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.

39. The kit of claim 37 wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin, dye particles and colloidal particles.

40. A diagnostic kit comprising at least two oligonucleotide primers, at least one of the oligonucleotide primers being specific for a DNA molecule according to claim 5.

41. A diagnostic kit according to claim 40, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA molecule according to claim 5.

42. A diagnostic kit comprising a at least two oligonucleotide primers, at least one of the primers being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

43. A diagnostic kit according to claim 42, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

45. A kit according to claim 44, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

46. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

47. A kit according to claim 46, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

48. A monoclonal antibody that binds to a polypeptide according to any of claims 1-4.

49. A polyclonal antibody that binds to a polypeptide according to any of claims 1-4.

50. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

51. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6 (SEQ ID NO: 99).

53. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO: 150).

54. A diagnostic kit comprising:

- (a) one or more fusion proteins according to any one of claims 50-53; and
- (b) a detection reagent.

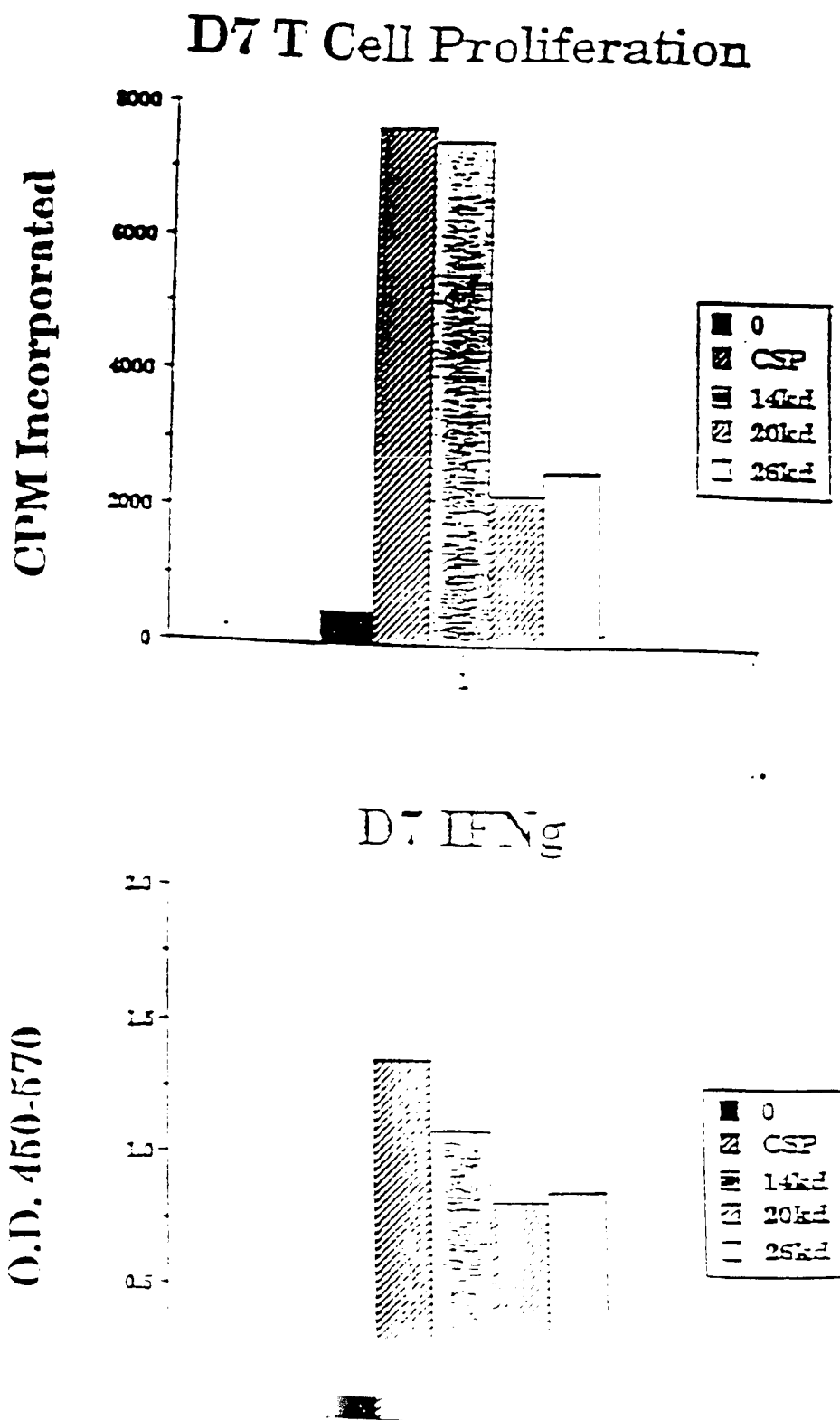
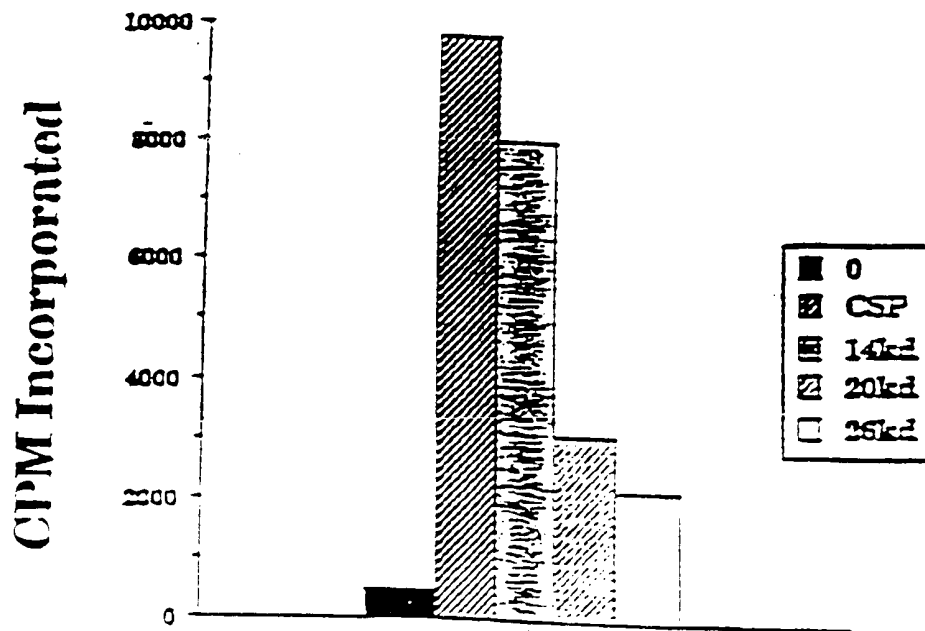
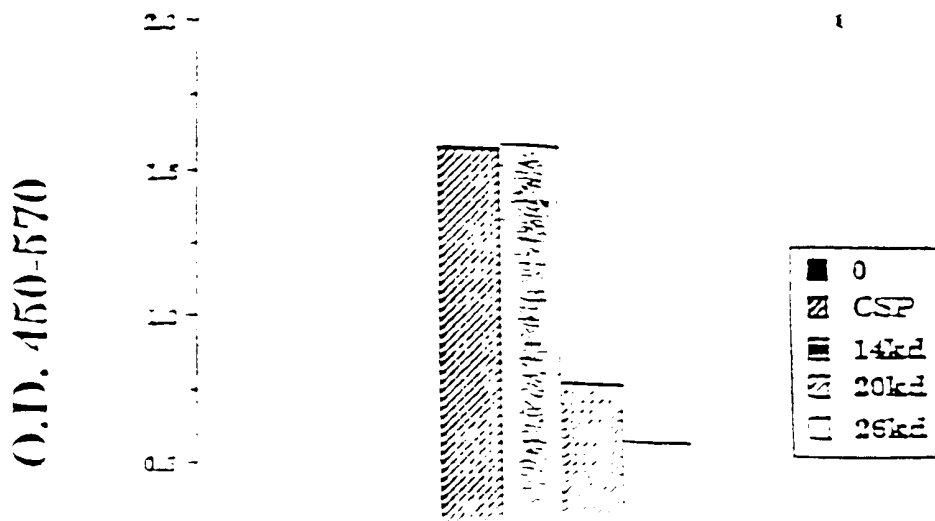


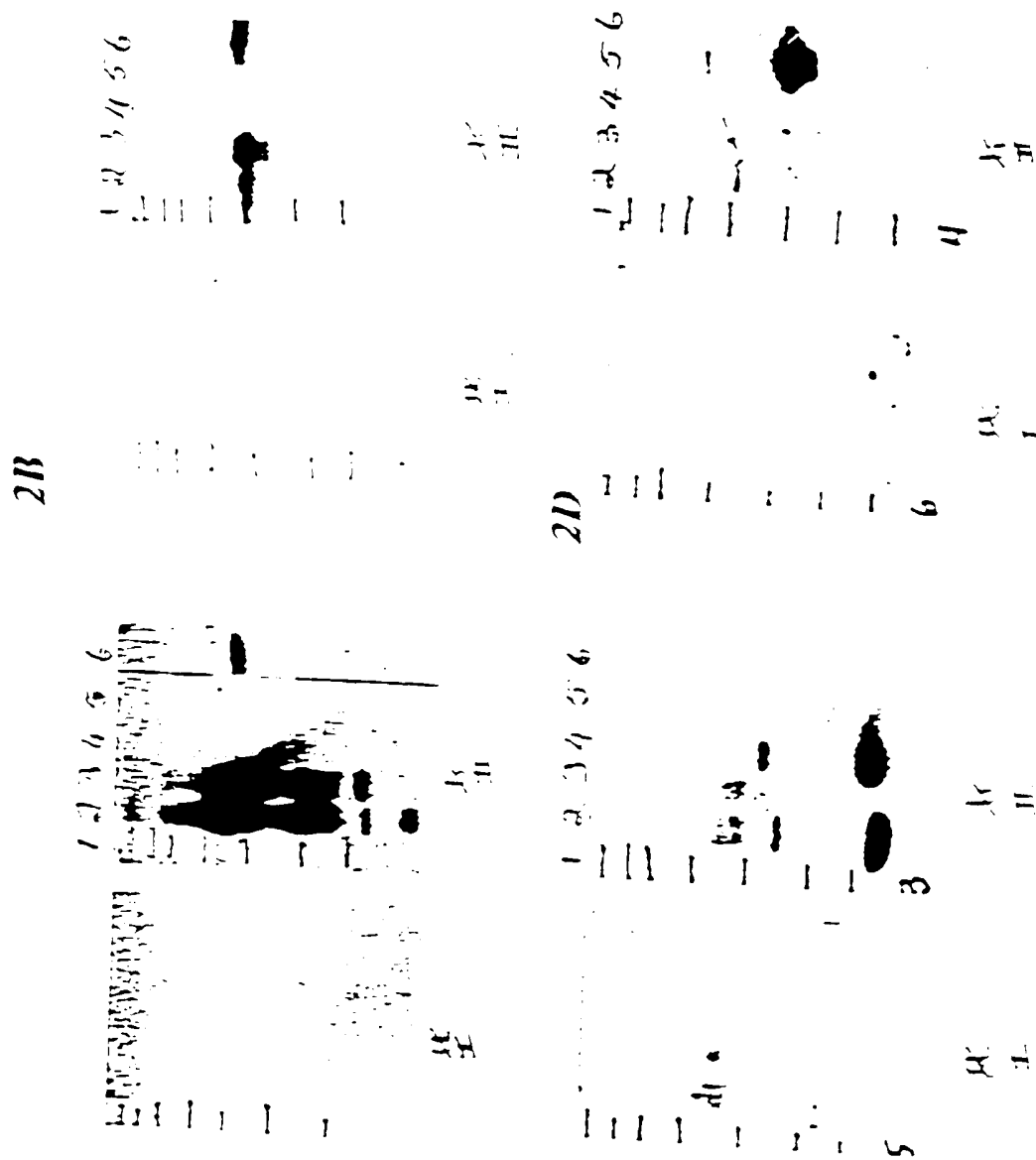
FIG. 1A

D160 T Cell Proliferation



D160 IFN_γ





FIGS. 2A-D

T cell clone 131TbH9 responds poorly to CSP

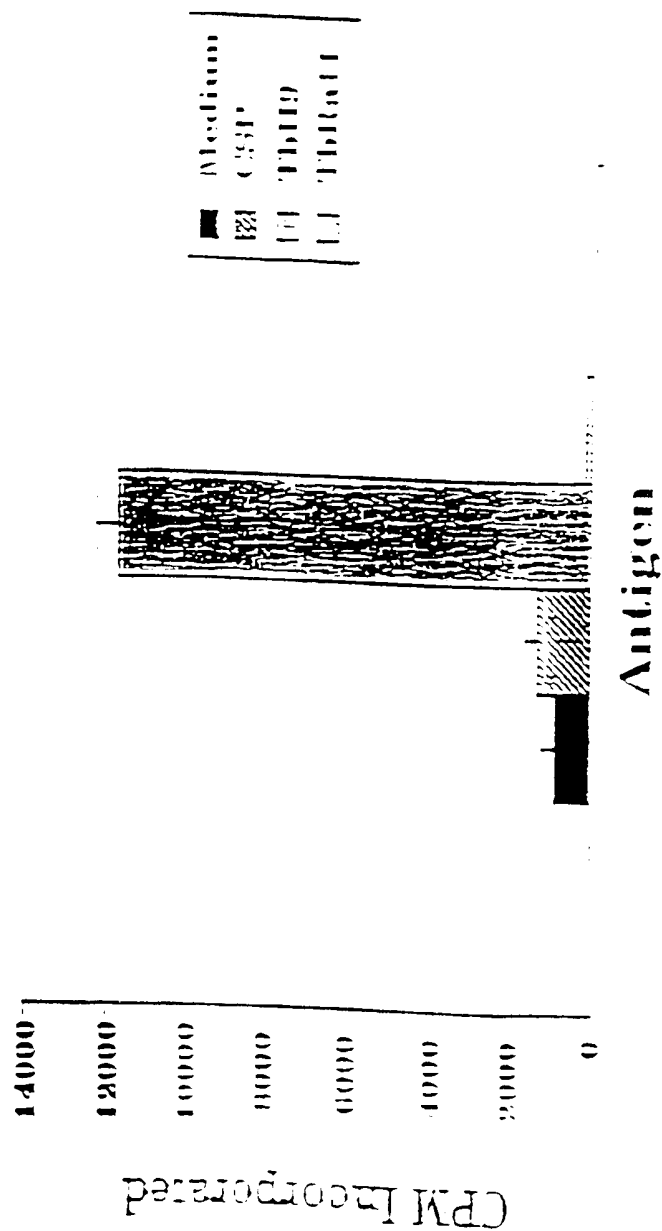


FIG. 3A

T Cell Clone PPD 800-10 IFN γ Production

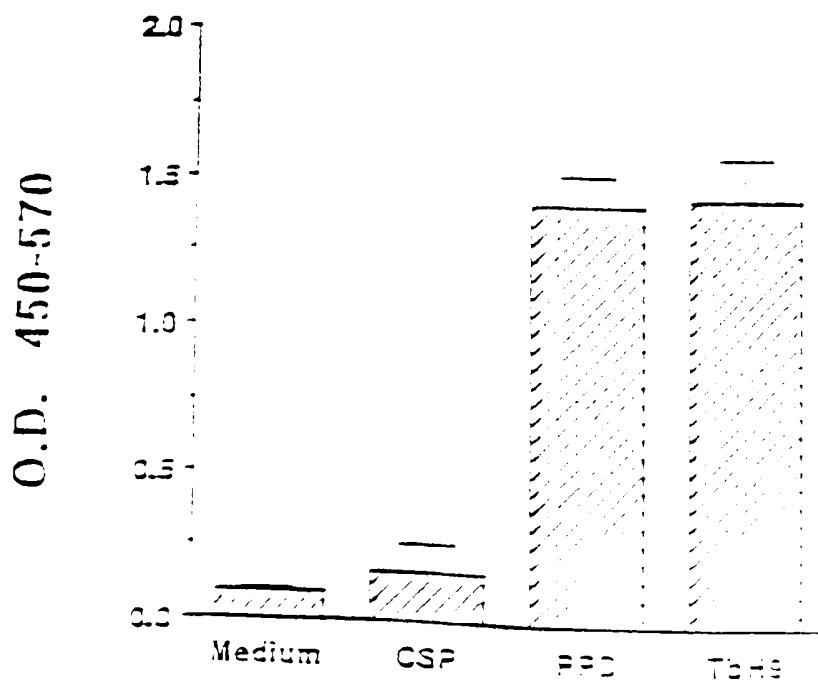


FIG. 3B

Anti-CD40L Antibody Titration

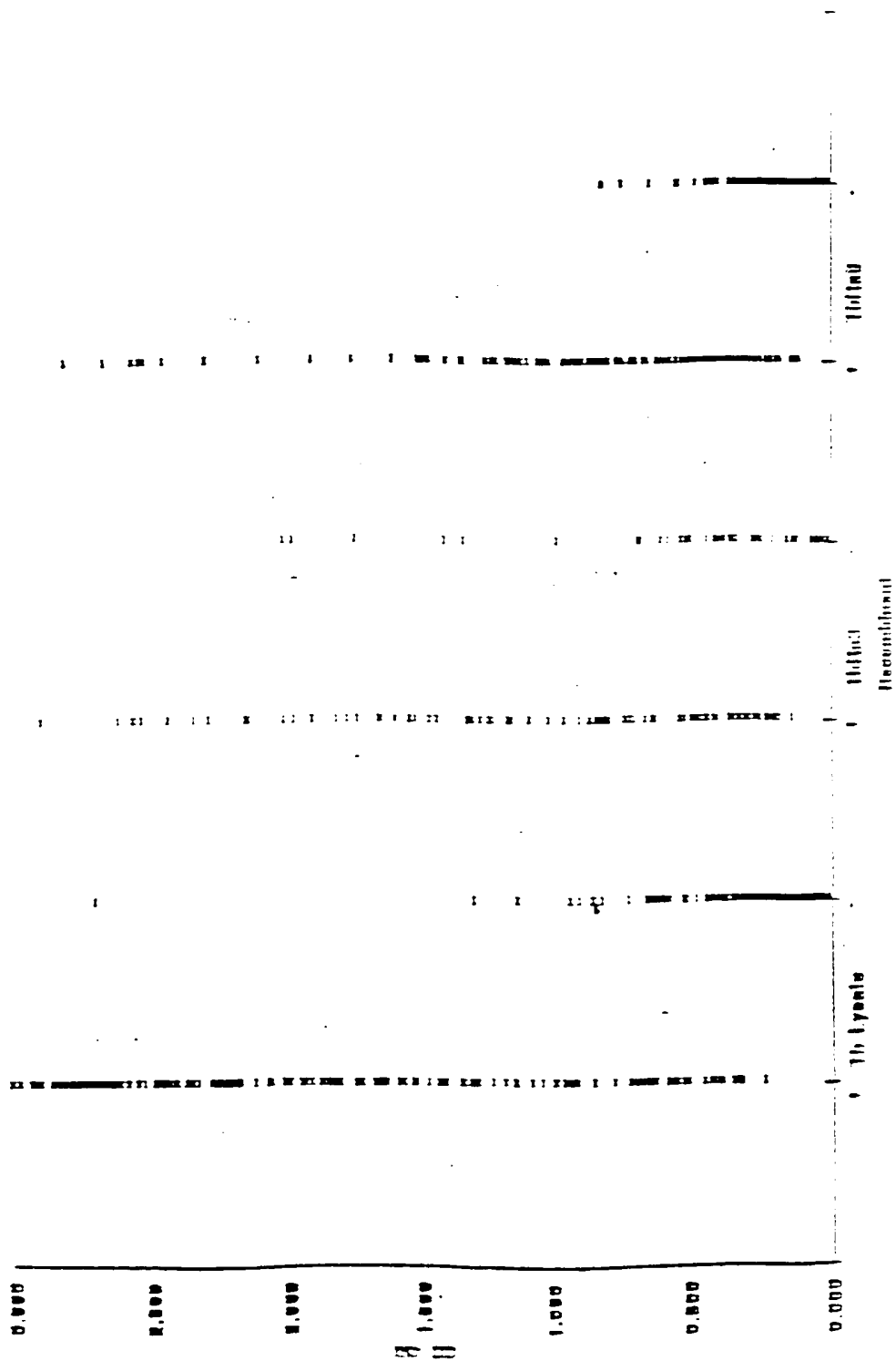


FIG. 4

FIG. 10. Sensitivity of Fluorescence and Peptides obtained using Human T0 Bone marrow procedure

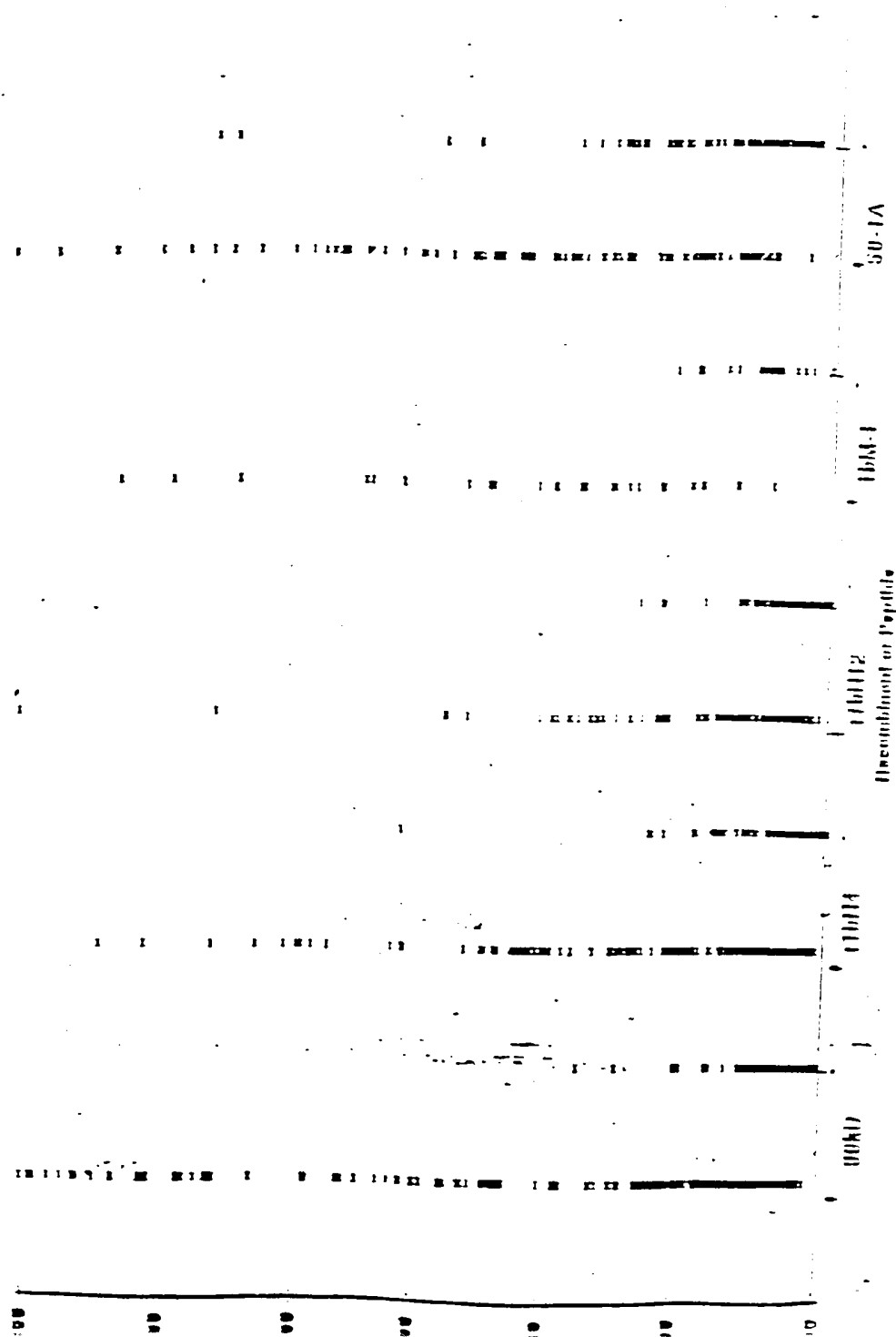


FIG. 6

Reactivity of Recombinant 30kD and 101kD antigens with sera from *M. tuberculosis* patients, PPD positive and normal donors

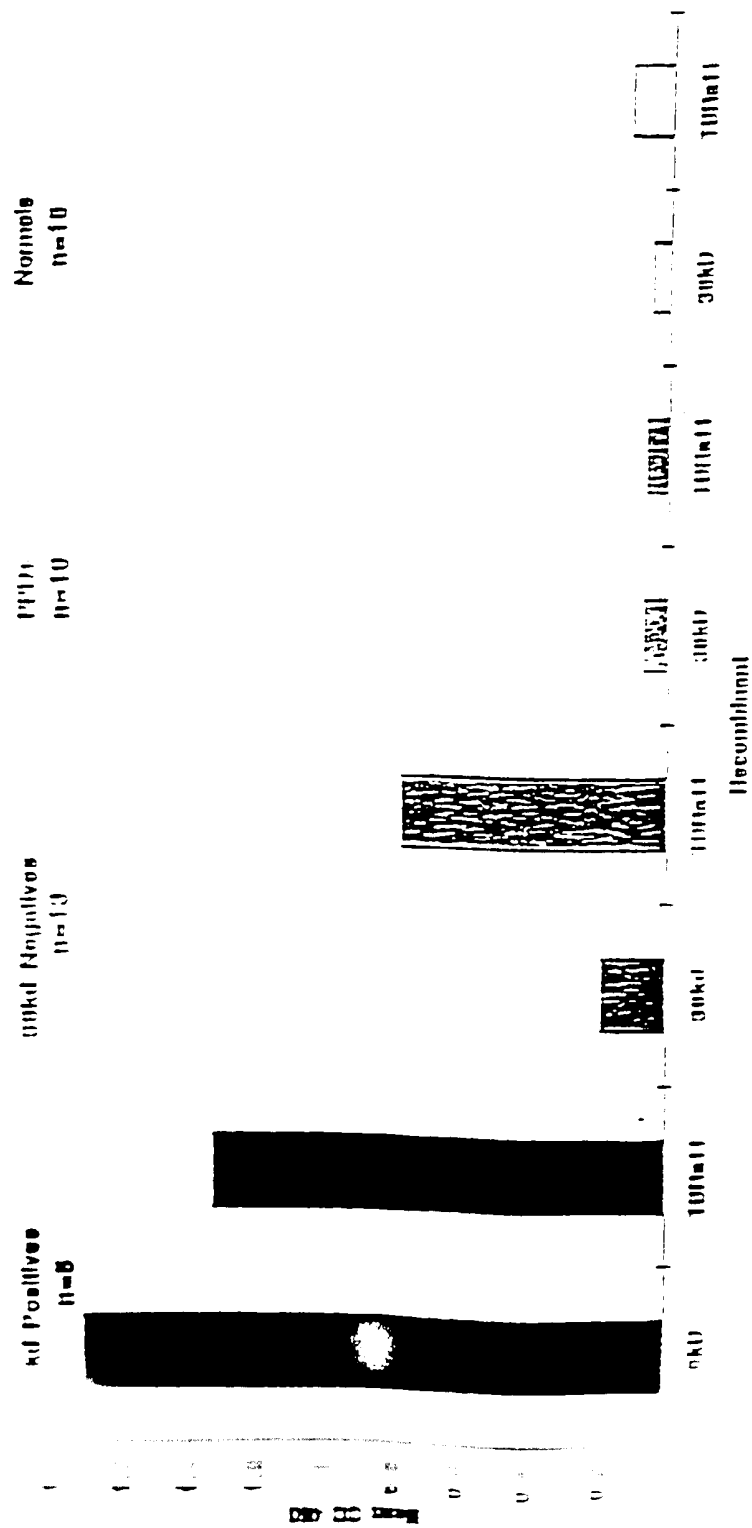
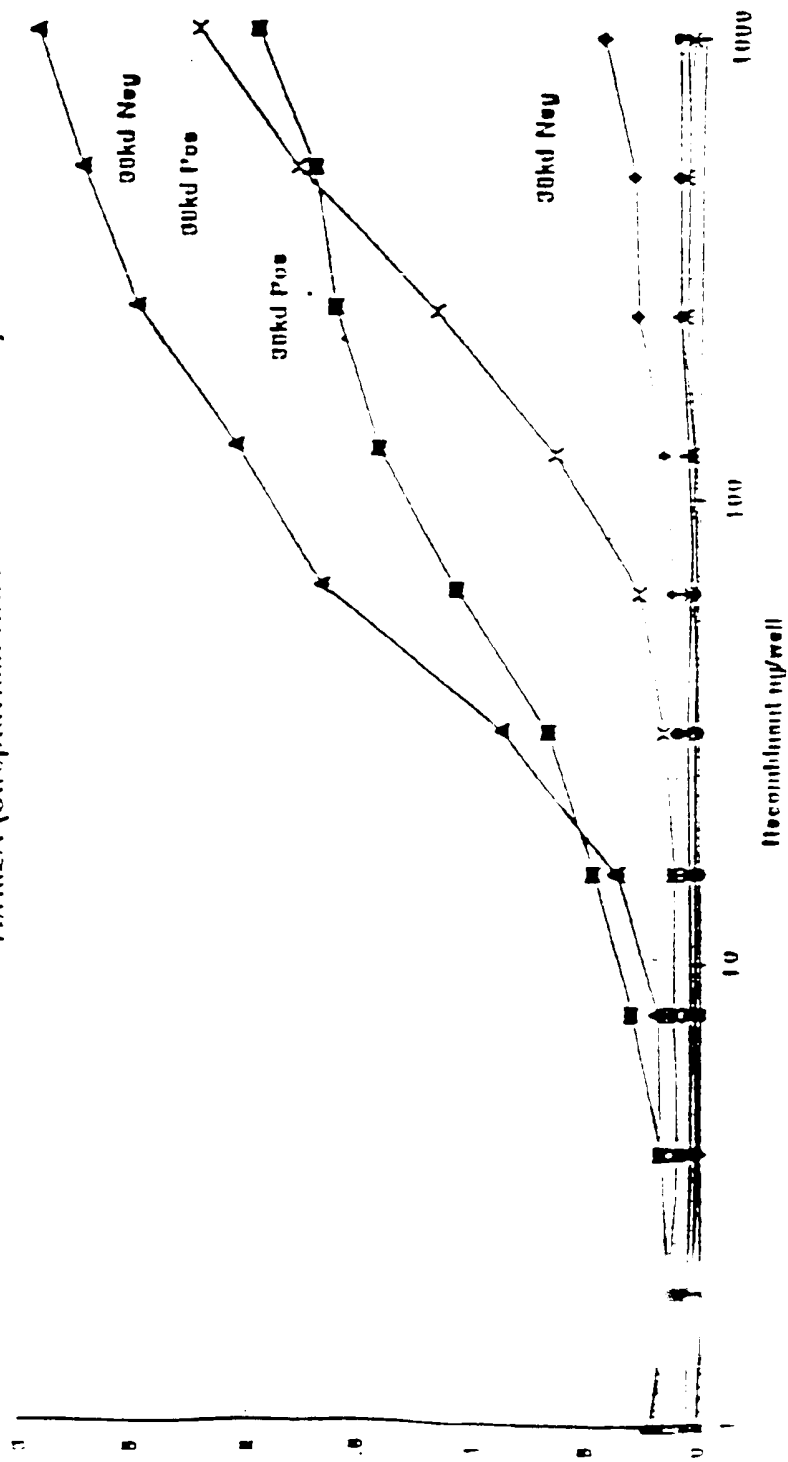


FIG. 7

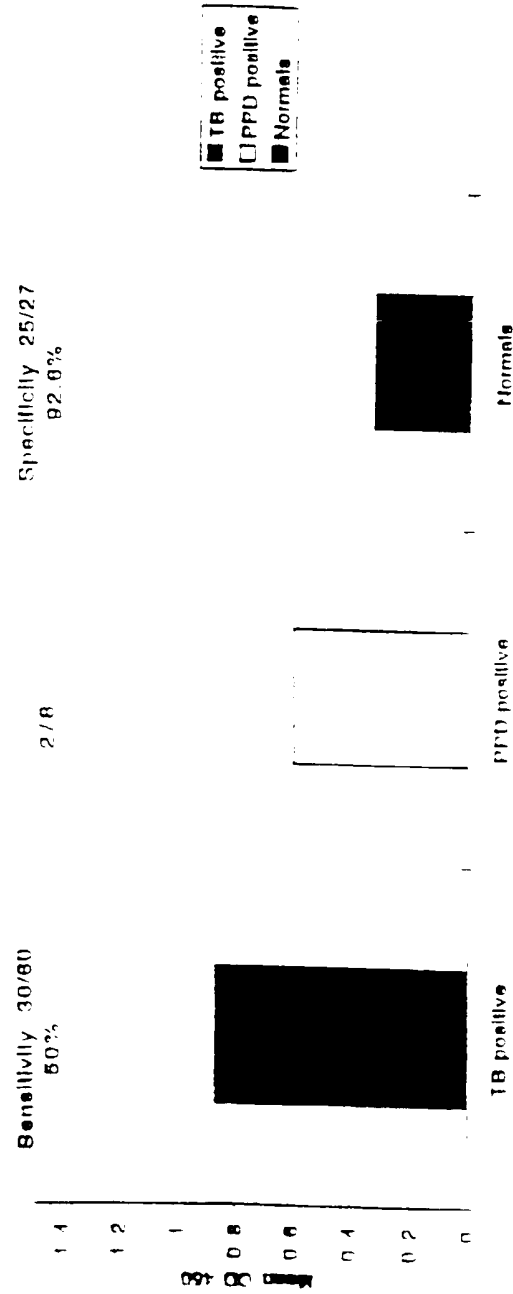
TbH2A (Streptococcus HRP/Protein A Blot)



10004
67004
60004
30004
A114
A110
A140
A141

FIG. 9

TbH29 ELISA reactivity in Streptavidin-HRP/Protein A-biotin system



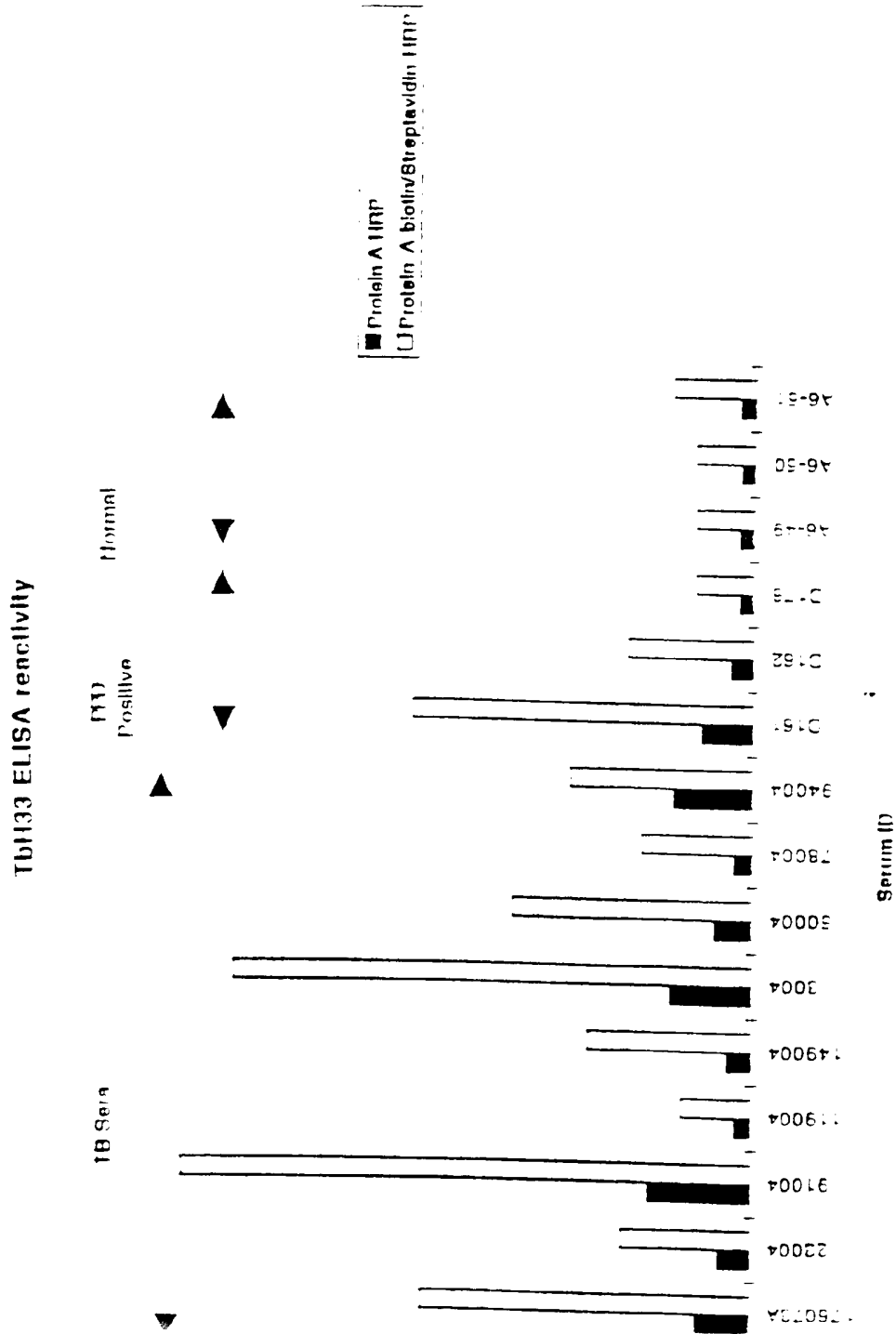
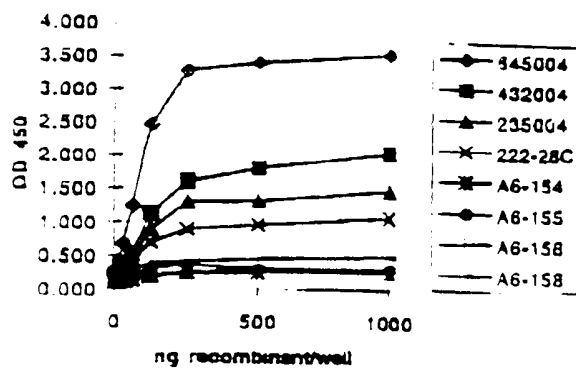
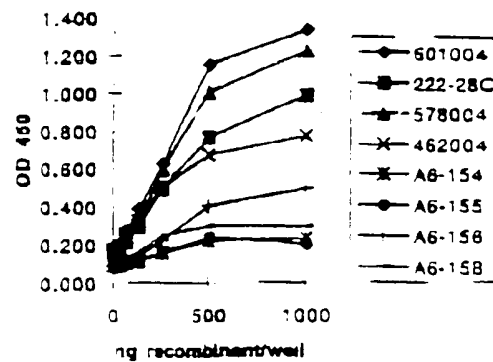


FIG. 10

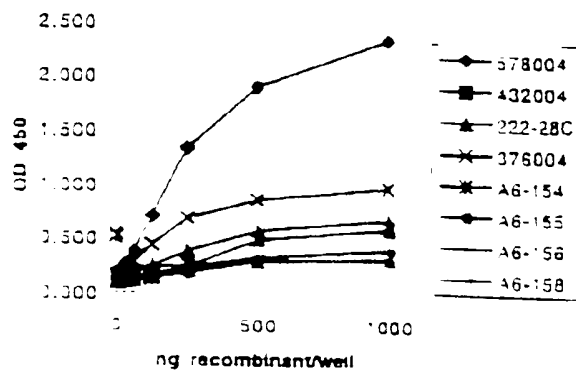
Titration of Mo-1 antigen with TB
positive and negative sera



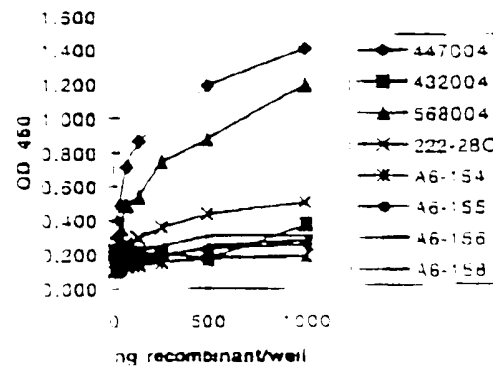
Titration of Mo-2 with TB
positive and negative sera



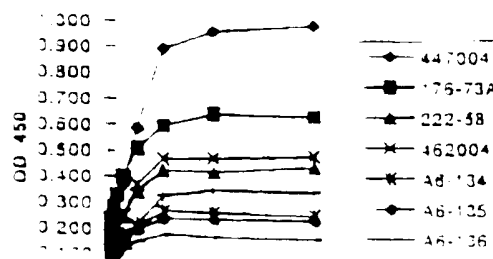
Titration of Mo-4 with TB positive
and negative sera



Titration of Mo-28 with TB
positive and negative sera



Titration of Mo-29 with TB
positive and negative sera



SEQUENCE LISTING

(1) GENERAL INFORMATION:

(A) APPLICANTS: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald

(B) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

(C) NUMBER OF SEQUENCES: 100

(2) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7090

(3) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.10

(4) IMPRINT APPLICATION DATA:

(A) APPLICATION NUMBER
(B) FILING DATE: 15-MAY-1998
(C) CLASSIFICATION:

(5) ATTORNEY/AGENT INFORMATION:

(A) NAME: Marki, David
(B) REGISTRATION NUMBER: 1,100
(C) REFERENCE NUMBER: 00101, 11001

(6) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 206 422 4400
(B) TELEFAX: 206 482-6731

(7) INFORMATION FOR SEQ. ID NO. 1

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG GTAGTTTGAA CCAACGCGAC AATCGACGGG CAAACGAACG CAAGAACACA	50
ACCATGAAGA TGGTGAAATC GATCGCCGCA GGTCTGACCG CCGCGGCTGC AATCGGCGCC	120
GCTGCGGCGG GTCTGACTTC GATCATGGCT GGCGGCGCGG TCGTATACCA GATGCAGCCG	180
GTCTCTTTGG GCGCGGCACT GCGCTTGGAC CCGGCATCCG CCGCTGACCT CCGGACCGCC	240
GCGCAGTTGA CGAGCCTGCT CAACAGCCTG GCGGATCCCA ACCTGTCTTT TCGGAACAAG	300
GCGAGTCTGG TCGAGGCGCG CATCGGGGCG ACCGAGGCGG GCATCGCGCA CCACAAGCTG	360
AACAAGCGCG CCGAGCACCG GATCTCTCGG GTCTGCTTCA GCTGACGAA GATCCAGCCG	420
GGGCGCGCGG GTTGGGCGAC CCGCGACCTT TCGCTCTCGG GTCGGAAGCT GTCTCGCGCG	480
GTGAGCGAGA ACCTCACTTT GGTGAATCAA GCGCGCTGGA TGCTGTCACG CCGATCGCGG	540
ATGAGGTTGG TCGAGGCGCG AGGCGAAGTG ATTGGCGGCG CCGNTTCAGC CCGCTCTTCA	600
GCTACCGCGG CCGCCTGCTG ACCGCTCCAT GTCGAAGACT CCGCGCTGTA GCACGCTGCG	660
GTCTCGCGAG GCGCGCGCGG ACCGCGCGCT GCAAGCGCTG GTGAGATAG GTGCTGCTG	720
GNCAGCAGNG ACGAGCGCGG NNTGCGCGCT TCTGCTGCTG GATGGA	780

C INFORMATION FOR SEQ ID NO:1:

1. SEQUENCE CHARACTERISTICS:
 - A. LENGTH: 789 base pairs
 - B. TYPE: nucleic acid
 - C. STRANDEDNESS: single
 - D. TOPOLOGY: linear

H. SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGATGATG ATCAAGTCA ATATGAATG AGTATAGAA GAACTGCTT GTTATCTTCA	120
CACTTCTTA TCGAGCTGGA GTTCTGCTT GAACTGCTT TCGAGAGTCT GTTCTCTTCA	240
GTGGAAGCGG TCGCGCGCGG GTTCTCTTCA GTTCTCTTCA AAGGAGGCGG CAACTGCTTCA	360
GTGCTCTTCA TACTGAGCA AGGCTATGAG TCGCTCTTCA GCGATCGCGA CAGGAGCTA	480
GTGCTCTTCA AGGCTATGAG TCGCTCTTCA GTTCTCTTCA GTTCTCTTCA AAGGAGGCGG	600
GTGCTCTTCA GTTCTCTTCA GTTCTCTTCA GTTCTCTTCA GTTCTCTTCA GTTCTCTTCA	720

GCGCGCGCGCT GGCGCGGATG TCGATCGGGG CGGTCTCTCG ACCTGCTACG ACCGGATTTT 540
 CGCTGATGTC CAGCATCTCC AAGATTGAT TCTTGGGAGG CTTGAGGGTC NGGGTGACCC 600
 CCGCGCGGGC CTCATTGCGG GGTNTCGGCN GGTTCACCC CTTACCNACT GCCNCCCGGN 660
 TTGCNAATTC NTTCTTCTCT GCGCNAAG GGACCTTAN CTTGCGCGCTN GAAANGGTNA 720
 TCCNGGGCCC NTCTTGAAN CCGCTCCCC CT 752

(2) INFORMATION FOR SEQ ID NO:3:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATATGCATC ACCATCACCA TCACATCTCT AACCGCCAG CGCTCGGGG GCGTCAGCA 60
 CCAACCGGACA CCGGGCCCCGA TCGATCTCTT AGCTTGAGTC TGGTCAGGCA TCGTCGTGAG 120
 CACCGCGATG CCGTATGTTT GTCTGTGATC CAGATATCGC GGCAATCCAA TCTCCCGCCT 180
 GCGCGCGGGC CTCCTGCAAA CTACTCGGGG AGGAATTTTC ACCTGCGCAT CAAGATCTTC 240
 ATCTGTGTCG CCGCTCTCTT TTTCTCTCTT TGTTCGGGTC TGGCCACGCG CGGCGCCAG 300
 AGCTACTGCG AGCAATTGAA AGGCACCAT ACCCGCCAGG CTGCGCAGAT TCAAAATCTCC 360
 CAGCGCTCTT ACAACATCAA CATCAGCTCT CCAATTTACT ACCCGCAGCA CAAGTCTCTT 420
 CAAAATTACA TCGCCGAGAC GCGCGACAAG TCGCTCAGCG CGGCGACATC CTCCTCTGCA 480
 TCGCAAGCGC CTACCAATT GAATATCAGT TCGGACAT ACCAGTCCCG CATACCTCTG 540
 CTCTTATCTT GATCTCTCTT CTCTTCTCTT TATCTCTT TCTCTCTCTT TCTCTCTCTT 600
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT 660
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT 720
 CAGCAAGCGC CAGCAAGCGC CAGCAAGCGC CAGCAAGCGC CAGCAAGCGC CAGCAAGCGC 780
 TCAAAATCTT ACAATCTCTT AATCACTCTT AATCACTCTT AATCACTCTT AATCACTCTT 840

(2) INFORMATION FOR SEQ ID NO:4:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGTATGAAC ACGGCCCGGT CCGATAACTT CCAGCTGTCC CAGGGTGGGC AGGGATTCCG	60
CATTCCGATC GGGCAGGCCA TGGCGATCGC GGGCCAGATC CGATCGGGTG GGGGGTCACC	120
CACCGTTCAT ATCGGGCCTA CCGCCTTCTT CGGCTTGGGT GTGTTCGACA ACAACGGCAA	180
CGGGCGACGA GTCCAACGCC TGGTCGGGAG CGCTCCGGCC GCAAGTCTCG GCATCTCCAC	240
TGGCGACGTG ATCACCAGCC TCGACGGGCG TCCGATCAAC TCGGCCACCG CGATGGCGGA	300
TGGCCTTAAC GGGCATCATC CCGGTGACGT CATCTGGTG AACTGGCAAA CCAAGTCGGG	360
CGGCACGGGT AAGGCGAACC TCACATTGGC CGAGGGACCC CCGGCCTGAT TTGCTCGYGC	420
ATACCAGCCG CCGCCCGGCC AATTGGA	447

(x) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 604 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTCCACTCC GGTGCGGAG CATCTCGCCC AGCAAAATGT TGGCAGCCCG CCAACGGAAT	60
CGGCTGATCC CAGGTGGAAG TTCTCGAAG CCGCCCGGCG CGAAGTATCC CTCGATCCCT	120
AAGCTGCGCA CCGCGAGCGT CGAATCGCC CGAGTGAGGA AGCGGCGAAT TTGCGCGGCG	180
CGGGCCACCG NGAGCGCGCG AATGGCGCGA CTGAGGAGGT GGNCACTCAT CCCCAGNCTG	240
CTCCAATCAA CCGTATTCCT CCGTGGCGG CATTTGACA ATTAAGTAA TGAGCGCAAA	300
CAATGATCG AAAAGCGGCG NGACGTCCT TTCTTTCTCT GTTATAGCT CATTGCTCG	360
CTTGGCGGCT ATCAGATCT CATTCTCTCA AACTGATCT CGAGCAACA TTCTTCGCG	420
CTTATGCTCA CCGTCTGCAAT CTCTTCTCT CTGCGCAAT CAGAGAGCG TTGATGNGA	480
CAAAAAGCTG CAGCAGGCGG AATTGNGGCG CCAANAAGC CTTATGNGCG CAGAGTNGCT	540
CTTCTCTCTCT CAGAGTNGCG CAGAGTNGCG CAGAGTNGCG CAGAGTNGCG CAGAGTNGCG	600

(A) LENGTH: 633 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANGTCG AACCACTCA CTAAAGGGAA CAAAAGCTNG AGCTCCACCG CGGTGGCGGC	60
CGCTCTAGAA CTAGTGKATM YYYCKGGCTG CAGSAATYCG GYACGAGCAT TAGGACAGTC	120
TAACGGTCCT GTTACGGTGA TCGAATGACC GACGACATCC TGCTGATCGA CACCGACGAA	180
CGGGTGCGAA CGCTCACCCT CAACCGGCGG CAGTCCCGYA ACCGCGCTCTC GCGCGCGCTA	240
CGGGATCGGT TTTTCCCGGY GTTGGYCGAC GCGGAGGYCG AGGAGGACAT CCACCTCCTC	300
ATCTTCACCG GYCGCGATCC GGTGTTCTCC GCGGACTGG ACCTCAAGCT AGCTGGCGCG	360
CGAGACCGCG CTGCGCGACA TCTACCGCG GTGCGCGCG ATGACCAAGC CGGTGATCGG	420
CGCGATCAAC GCGCGCGCG TCACCGCGCG GTTCCAACTG GCGCTGTACT GCGACATCCT	480
GATCGCGTCC GAGCAGCGCG GCTTCGCGCA CACCCACCGC CGGGTGGGGC TGCTGCGCAC	540
CTGGGGAATC AGTGTGTGCT TCGCGCAAAA GGTGCGCATC GGNCTGGGCT GGTGGATGAG	600
CGTGACCGCG GACTAGCTGT CGGTGACCGA CGC	633

2. INFORMATION FOR SEQ ID NO:7:

A. SEQUENCE CHARACTERISTICS:

A. LENGTH: 1362 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:7:

TAATATGA TGTCTGCA TATGCTGTA TATATGAT TATCTGCT TGGGAGAT	60
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	120
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	180
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	240
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	300
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	360
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	420
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	480
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	540
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	600
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	660
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	720
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	780
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	840
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	900
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	960
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1020
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1080
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1140
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1200
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1260
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1320
TAATATGA TATATGTA TATATGTA TATATGTA TATATGTA TATATGTA	1362

CGCGCGCGGAC CGTTCGGGCC GGATGTGGG GCGGAATACC TGGGCACCGC GGTGCAATTC	540
CAGTTCATCG CACGCTGGT CTTGGTGCTG CTGGACGAAA CTTTCCTGCC GGGGGGCCCC	600
CGCGCGCAAC AGCTCATGCC CGCGCGCGGT GGACTGGTGT TCGCCCGCAA GGTGCGCGCG	660
GAGCATCGGC CGGGCCGCTC CACCGCGCGG CTGAGCGCG GAACGCTGCC CGACGATCTG	720
CGATGGGCAA CACGCTCGA GCGCATAGCA ACCGCGTTCC CCGCGCTCAG CCACCACTG	780
GACACCGCGC CGCACCTGCC GCGACCGACT CGTCAGGTGG TCAGGCGGGT CGTGGGGTCC	840
TGGCACGGCG AGCCAATGCC GATCAGCAGT CGCTGGACGA ACGAGCACAC CGCGAGCTG	900
CGCGCGGACC TGCACGCGC CACCGCTCTT GCGCTGCTGA CGGCGCTGCC CCGCATCAG	960
GTACCGSACC ACGACGTGCC CGCGCGCGCA TCGCTGCTG ACACCGATGC GCGCGTGGT	1020
CGCGCGCTCG CTTGGGGCGC TTACACCGC CGCGCGCGCA TGGCACCTG GATCGCGCG	1080
CGCGCGGAGG GCGAGGTGT CGCGCGAAC CGACTGGGT GAGTGTGCC GCGCTGTGCG	1140
TAGGTGTCA TCGCTGGCC GAGGATCTC CGGCGCGCA ACGAGGTGG CGACACAGGT	1200
GGAAGCTCG CCGACTGGCT TCGCGCGCA CCGCTGCTG GCGCTTCGCT TGGCGCGACT	1260
CGCGCATCAG CTGGCGCGC CGCTTGGC GAAGGTCCAG CTACCGTGC CCGACCGAA	1320
CGCGCGGAGG CTCACCGCG GTCACCGTGC CGCGCGAAGG AA	1360

1. INFORMATION FOR SEQ ID NO. 1:

(a) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1458 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(b) SEQUENCE DESCRIPTION (Seq. ID No. 1)

CGAGCAATTC CGATATGCC CGCACTATC AAAATTC TCGCATCA ATGCTGCT	1
CTATGCTGCT GTTGAGGAC ATTGAGATC TGTGAGC TGTCTGCGT GAACTGCT	120
TGATGAGCT CGCGCTCTT TACATCATC ATGCGAGCG CGCGCGGAG CTGCGGACCG	180
CTACCGCTT TGTGCGCTG TGGACGACT AAAATTCG TTTGGCGCG CTGATGCTA	240
TATGATGCT TATGCTGCT TCGACGACT AAAATTCG TTTGGCGCG CTGATGCTA	300

CGATTGAGGA TTGCTGCAA TCGATCTTTG TGACCTGGG ACAGGCCGCC GAGCTGCAGC 540
 GGGCTGGAGG CGGCACCGGA TATGCGTTCA GGCACCTGCG ACCCGCCGGG GATCGGGTGG 600
 CCTCCACGGG CGGCACGGCC AGCGGACCGG TGTCGTTTCT ACGGCTGTAT GACAGTGCCG 660
 CGGGTGTGGT CTCCATGGGC CGTCGCCGGC GTGGCGCCTG TATGGCTGTG CTTGATGTGT 720
 CGCACCCGGA TATGTGTGAT TTCGTCACCG CCAAGGCCGA ATCCCCCAGC GAGCTCCCCC 780
 ATTTCAACCT ATCGGTTGGT GTGACCGACG CGTTCCTGCG GGCCGTGCGA CGCAACGGCC 840
 TACACCGGCT GGTCAATCCG CGAACCGGCA AGATCGTCGC GCGGATGCCC GCCGCCGAGC 900
 TGTTCGACCG CATCTGCAAA GCGCGCCACG CGCGTGGCGA TCCCGGGCTG GTGTTTCTCG 960
 AGACGATCAA TAGGSCAAAC CGCGTCCCGG GGAGAGGGCG GATCGAGGCG ACCAACCCCT 1020
 CGCGCGAGGT CCGACTGCTG CTTACGAGT CATGTAATCT CGGCTCGATC AACCTCGCCC 1080
 CGATGCTCCG CGACGCTCGG GTGACTGCGG ACCGCTCGA GGAGGTGCGG GTGTGCGCG 1140
 TGCGTTCTCT TGATGACGTC ATGATGTCA CGCGCTACCC CTTCCCGGAA CTGGGTGAGG 1200
 TGGCGCGCGG CACCGCGAAG ATCGGGCTCG GAGTCATGGG TTGCGCGGAA CTGCTTCCCG 1260
 CATGCGGTAT TCCGTACGAC AGTGAAGAAG CGGTGCGGTT AGCCACCGCG CTCATGCGTC 1320
 CTATACAGCA CGCGCGCGAC ACGGCATCGG CGACGCTCGG CGAAGAGCGG CGCGCATTCG 1380
 TGTGTTGAG CATACCGCGG TCGCGCTGCT CGCGCGCGAG CGCGAAGCGA GAGTGAAGT 1440
 TGTGCTGCTG TACGCGCA 1450

1. INFORMATION FOR SEQ ID NO. 1:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 562 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO. 1:

AGGTTTAAAT GTGCTGGAT CTGGAACCGG TGTGCGCTT AGGTACCGAG ATTAACTGGT 50
 TGTGACCGG GTTGTGCTG CGCATCGCG TGTGTTAACT CGCGATCGCG GTGCTATTCG 100
 TGTGTTAACT GTGACAGG AGCGCGCTT CGAAGCTGCT GAGTGGCGA AAACCGCGCT 150

CCGCGGTGCA CCGCGCGCCG GTGCTCAAGG AAGGGGACGA TTGCCCCGAT TCGACGCTGG	360
CGCTCAAAGG TTTGACCAAC GCGCCGCGAGT ACTACGTCGG CGACCAGCCG AAGTTCACCA	420
TGGTGGTCAC CAACATCGGC CTGGTGTCTT GTAAACGCGA CGTTGGGGCC GCGGTGTTGG	480
CCGCCTACGT TTA CTGCTG GACAACAAGC GGTGTGTGTC CAACCTGGAC TCGCGCCCTT	540
CGPATGAGAC GCTGGTCAAG ACGTTTTCCT CCGGTGAGCA GGTAACGACC GCGGTGACCT	600
GGACCGGGAT GGGATCGGC CCGCGCTGCC CATTGCCCCG GCGGCGGATC GGGCCGGGCA	660
CGTACAATCT CGTGGTACAA CTGGGCAATC TGGCTGCTT GCGGGTTCCG TTCATCTGA	720
ATCAGCCGCG CCGCGCGCCG GGGCCGGTAC CCGCTCGGGT TCCAGCGCAG GCGCCTCCGC	780
CGAGTCTCT CCGCGCAAGG CGATAATTAT TGATGCTTGA TGGTCGATTC CGCCAGCTCT	840
GACAACCCCT CGCTCTCTCT	862

10 INFORMATION FOR SEQ ID NO:10:

(1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 622 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

11 SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTATGAGCA CCGCGCAAGG GTGCTCAAGG AAGGGGACGA TTGCCCCGAT TCGACGCTGG	60
GAGTCAAAGG TTTGACCAAC GCGCCGCGAGT ACTACGTCGG CGACCAGCCG AAGTTCACCA	120
TGGTGGTCAC CAACATCGGC CTGGTGTCTT GTAAACGCGA CGTTGGGGCC GCGGTGTTGG	180
CCGCCTACGT TTA CTGCTG GACAACAAGC GGTGTGTGTC CAACCTGGAC TCGCGCCCTT	240
CGPATGAGAC GCTGGTCAAG ACGTTTTCCT CCGGTGAGCA GGTAACGACC GCGGTGACCT	300
GGACCGGGAT GGGATCGGC CCGCGCTGCC CATTGCCCCG GCGGCGGATC GGGCCGGGCA	360
CGTACAATCT CGTGGTACAA CTGGGCAATC TGGCTGCTT GCGGGTTCCG TTCATCTGA	420
ATCAGCCGCG CCGCGCGCCG GGGCCGGTAC CCGCTCGGGT TCCAGCGCAG GCGCCTCCGC	480
CGAGTCTCT CCGCGCAAGG CGATAATTAT TGATGCTTGA TGGTCGATTC CGCCAGCTCT	540
GACAACCCCT CGCTCTCTCT	600

(A) LENGTH: 1200 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

GGCGCAGCGG TAAGCCTGTT GGCCGCCGGC AACTGGTGT TGACAGCATG CGGCGGTGGC      60
ACCAACAGCT CGTCGTCAGG CCGAGGCCGA ACGTCTGGGT CGGTGCACTG CCGCGGCCAAG      120
AAGGAGCTCC ACTCCAGCGG CTCGACCCCA CAAGAAAATG CCATGGAGCA GTTCGTCTAT      180
CGCTACGTGC GATCGTGCCG GGGGTACAGC TTGGACTACA ACGCCAACGG CTCGCGTGCC      240
GGGTGACCCG AGTTTCTCAA CAACGAAACC GATTTCCGCC GCTCGGATGT TCCGTTGAAT      300
CGCTCGACCC CTCACCTGA CCGGTCCGCC CAGCGGTGCC GTTCCCGGGC ATGGGAGCTG      360
CCGACGGTGT TCGGCCCGAT CGCGATCACC TACAATATCA AGGCGGTGAG CACGCTGAAT      420
CTTCACCGAC CCACTACCCG CAAGATTTTC AACGCCACCA TCACCGTGTG GAATGATCCA      480
CAGATCCAAG CCGTCAACTC CGGCACCGAC CTGCCGCCAA CACCGATTAG CGTTATCTTC      540
CGCAGCCACA AGTCCGGTAC CTCGGACAAC TTCCAGAAAT ACCTCGACCG TGTATCCAACT      600
GGGGCGTGGG GCAAAGGCCG CAGCGAAACG TTCAGCGGGG CCGTCCGGCT CGGCCCCAGC      660
GGGAACAACG AACCTCCGC CTTACTGCAG AGGACCGACG GGTCCATCAC CTACAACGAG      720
CGCTGCTTTG CGTCCGTAA CCACTTAACT ATCTCCGAGA TCATCACCTC CGCGGGTCCG      780
GATTAATGTC CGATCACCAC CGAGTCCGTC GTTAAGACAA TCCCGCGGGC CAAGATCATG      840
CGATAAGGCA ACGACCTCGT ATTCGACAGC TCGTCTGTTT ACAGACCCAC CCAGCCTGGC      900
CTTTACGCGA TCGTCTCGCC GACCTATGAG ATTCTCTGTT GAAATACCC CGATCCGAGC      960
CGCTATCTTC GTTAAGGTA GTTATGTA ATCTGATGTT GTTAAGGTA GAAAGGCTAT      1020
CGCTATCTTC GTTATGTA GTTATGTA ATCTGATGTT GTTAAGGTA GAAAGGCTAT      1080
GATTAATGTC CGATCACCAC CGAGTCCGTC GTTAAGACAA TCCCGCGGGC CAAGATCATG      1140
CGCTGCTTTG CGTCCGTAA CCACTTAACT ATCTCCGAGA TCATCACCTC CGCGGGTCCG      1200
  
```

SEQUENCE INFORMATION FOR SEQ ID NO: 11:

SEQUENCE: 1200 base pairs

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGCT GCAGGTCGTG CTGTTGACG AACTGGGCAT GCCGAAGACC AAACGCACCA	50
AGACCGGCTA CACCACGGAT GCCGACGGCC TGCAGTCGTT GTTCGACAAG ACCGGGCATC	120
CGTTTCTGCA ACATCTGCTC GCCCACCAGC ACGTCACCCG GCTCAAGGTC ACCGTCGACG	180
GOTTGCTCCA AGCGGTGGCC GCCGACGGCC GCATCCACAC CACGTTCAAC CAGACGATCG	240
CCCCGACCGG CCGGCTCTCC TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGCACCG	300
ACCCCGCCCG CCGGATCCCG GACCGCTTCC TCGTCGGGGA CCGTTACGGC GAGTTGATGA	360
CGCCCCACTA CAGCCAGATC GAGATCGCGA TCATCGGGCA CTTGTCCGGG GACGAGGGCC	420
TCATCCAGGC GTTCAACACC GCGGAGGACC TGTATTGCTT CTTCCGCTCC CCGGTGTTCC	480
CTGTCCCATC CGACGAGGTC ACCGCGGAGT TCGCGCCCGC GGTCAAGGCC ATGTCTACG	540
GGTGTGTTTA CCGGTTGAGC GCCTAGGGCC TCTCCAGCA GTTGAAAATC TCCACCGAGC	600
AAGTCAACGA GCAGATGGAC GCGTATTTCG TCGGATTCCG CCGGGTCCGC GACTACCTGC	660
CGCGCTTACT CGAGCCGGCC CGCAGGAGC GCTACACCTC CACGGTGTCT GCGCGTCCGC	720
GCTACCTGTC CGAGCTGGAC AGCAGCAACC GTCAAGTCCG GGAGGCCGCC GAGCGGGCCG	780
CGCTGAACCT CGGATCCAG CGCAGCCCGC TCGACATCAT CAAAGGTGCC ATGATCCAGC	840
TGGAAAGGCT GTCGAACGAG GACAGCTCCG CTCTCCCATC CTGCTCCAG CTCGACGAGC	900
AGCTGCTGTC TGAATGCGC TCGGTCGAGC GCGAGCGCGT CGAGGCCCTC CTCGCGGACA	960
AGATGCGCGC TCTTACGCG CTCGACCTCC CTCTGAGGCT CTCGCTGCGC TACCGCGGCA	1020
GCTGCGAGCT TCGGCGGAG TCACTGCGCA CTCTGATCT CGGCGCGGAA CTCGCGGATT	1080
CTCTGAGGCT GATTTAGGCT TCTGAGGCT TCTGAGGCT TCTGAGGCT TCTGAGGCT	1140
TCTGAGGCT TCTGAGGCT	1150

INFORMATION FOR SEQ ID NO:12

SEQUENCE CHARACTERISTICS

- A LENGTH 1151 base pairs
- B TYPE nucleic acid
- C STRANDEDNESS single
- D TOPOLOGY linear

TGGGGGCTCG GGTTCGGCAT GGTCAAACAG GTGGTGGTCA ACCACGGCGG ATTGGCTGCGC 120
ATCGAAGACA CCGACCCAGG CGGCCAGCCC CCGGAACGT CGATTTACGT GGTGGTCCCC 180
GGCCGTGGA TGGCGATTCC GCAGCTTCCC GGTGCGACGG GTGGCGCTCG GAGCACGGAC 240
ATCGAGAACT CTCGGGGTTC GGCGAACGTT ATCTCAGTGG AATCTCAGTC CACGCGCGCA 300
ACCIAGTTGT GCAGTTACTG TTGAAAGCCA CACCCATGCC AGTCCACGCA TGGCCAAGTT 360
GGCCCGAGTA GTGGGCCTAG TACAGGAAGA GCAACCTAGC GACATGACGA ATCAGCCACG 420
GTATTGGCCA CCGCCGCAGC AGCCGGGAAC CCCAGGTTAT GCTCAGGGGC AGCAGCAAAC 480
GTACAGCCAG CAGTTGCACT GGCCTTACCC ACCGTGCGCG CCGCCGCAGC CAACCCAGTA 540
CGCTCAACCC TACGAGGCGT TGGGTGGTAC CGCGCGCGCT GTGATAGCTG GGTGATTCC 600
GACCATGAGC CGCGCTGCTG GATGGTTCC GCAACCGCTT GTGCGAGGCA TGTTCGGCAT 660
CGCGCGGCTG ACATAGCGCG TGGTGTCCCG CGCATGCGCG CGCGCGGCGG CATCGCTGGT 720
CGGGTTCAAC CGGGCAGCGG CGGGCGCCAG CGCGCGGCGA GTGGGTGCGA GCGCGCGCGC 780
AAGCATCCCC GCAGCAAACA TGGCGCGGG GTGGGTGCGA CAGGTGCGCG CCAAGGTGGT 840
CGCGAGTGTG GTCATGTTGG AAACGATCT CGCGCGCGAG TCGGAGGAGG GTCGCGGAT 900
CATTTGTGTG GCGGAGGGGC TGATGTTGAC CAACAACGAC GTGATGCGCG CGCGCGCGAA 960
CGCTGCGCTG CGCAGTCCCG CGCGGAAAAC GAGGTAAAG TTCTGTGAGG CGCGGAGCGC 1020
ACCGTTGAGG GTGGTGGGGG GTGACCGGAG TATGATATG GCGGTGCTCG GTGTTCAGCG 1080
GTGTGCGGG GTGACCGCGA TGTGCTGGG TTCTGCTCG GAGGTGAGG TCGGTGAGCG 1140
GTGTGCGGG ATCGGCTCGC CGCTCGGTTT GAGGCGGAG GTGACCGCGG GTATGCTGAG 1200
GTGTGCGAG GTTCACTGT GAGGAGCGCG GAGGCGGAG AAGGAGAAAG GTGTGCTGAG 1260
GTGTGCGAG AGGAGCGCGG GTATGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1320
GTGTGCGAG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1380
GTGTGAGAG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1440
GTGTGAGAG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1500
GTGTGAGAG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1560
GTGTGAGAG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG GTGTGAGCG 1620

GTGATGAAGG TCGCCGCGCA GTGTTCAAAG C

1571

(12) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:14.

CTCCACCCCG GTGCGCGCGG GTGTAGAAGT AGTGGATCCC CCGGGCTGCA GGAATTGCGG	60
ACGAGGATCC GACGTCCGAG GTTGTCGAAC CCGCCGCGCG GGAAGTATCG GTCCATGCGT	120
AGCCCGCGCA CCGCGCGCGG CGGAATGCGG TGAGTGAGGA GCGCGCGCAAT TTGCGCGCGG	180
CGCGCGCGCG CGAGCGCGCG AATGCGCGCA GTGAGGAGCG GCGCAGTCAT GCGCAGCGTG	240
ATCCAATCAA CCTGCATTCC GCGTCCGCGG CCATTTCACA ATCGAGGTAG TGAGCGCAAA	300
TGAATGATGG AAAACGCGCG GTGACGTCCG GTGTTGCGGT GGTGCTAGGT GCGTGCCTGG	360
GCTTGTGGGT ATCAGGATGT TCTTCGCGCA AACGTGATGC CGAGGAACAG GGTGTTCCCG	420
TGAGCGCGAC GCGTCCGAG CCGCGCGCTC TCGCGAGAT CAGGCAGTCC GTTGATCCGA	480
CAAAAGGCTT GACGAGCGTG CAGGTAGCGG TCGGAACGAC CGGAAAGTC GACAGCTTGC	540
TGGTATTAG CAGTCCGAT TCGACGTCC GCGCGAATGT GGTGCGCGA AAGGCGCTAT	600
CGAGGTAGCA CAGGAGGAG GTGTGCGGT TTGCGGTACA AGCGAGAAC ATCTCGGTGA	660
CACTGTTGCA CCACTGGAGC AATGCGCGT CGATTTCGTA ACTGTCAACT TCGCGCGTGC	720
TGGATGCTGC GCGTGGGTG AAGCAGCTGC TGTGCGGTGT TCGAAGCTG CAGCGCGAAG	780
TTGTTAAGT ATAGAGTGA ATGTTAAGT TGAAGTGA ATGTTAAGT ATGTTAAGT	840
TTGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT	900
ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT	960
ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT	1020
ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT ATGTTAAGT	1058

SEQUENCE INFORMATION FOR SEQ ID NO:14

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC	60
GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCGC CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTGTTT AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCAGG GTTTGCGGTT GATTCGTGCG ATTTTGTGTC TGCTCGCGGA GGCCTACCAG	300
CGCGGGCCCA GGTCCGCGTG CTGCCGTATC CAGGCGTGCA TCGCGATTCC GCGCGCCAGC	360
CGCGAGTTAA TGCTTCGCGT CGACCCGAAC TGGCGGATCC GCGCGGAGC TGATCGATGA	420
CGCTGGCCAG CGCTCGATC CGCGAGTTCC CGAGGAAAC GTGCTGCCAG CGCGGTAGGA	480
AGCGTCCGTA GCGCGCGGTG CTGACCGGCT CTGCTGCCG CCTCAGTGCG GCGAGCGAGC	540
GC	542

(D) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 918 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AGTTGCGGCG CGCGGCTTCC TTCCCGCAT TCGCGCGGTC CGCGATCAGC TGCGCATCGC	60
AGCATCAGC CGCTTTGCGG CGCGCAGGCG CGCTGCGGCG GCGCGCGCGG ATGCCACCGC	120
CTCAAGCTTGG CGCGCGCGCG CGCGATTTTC ATCAAGAGCT CGCGCGCGCG AGCGCTTAGC	180
CGCTTTGCGA CGCTGCGGCG CGCTGCGGTT TCGCGGCGG GAGCGCGAAT GAACCGCGCG	240
GAAGCGCGCG CGCGGCGGCG TTCCCGGCTT TCGCGGCGG GCGCGCGCGG CGCGCAATTG	300
CGTAAGAGCG AMGAGCGGTT CGCGCGAGCG CTGCGGCGCT TAACGGCGCT CGCGCGCGCG	360
CGCGCGCGAG CGCGGATTAC CGCGGTTGCG CTGCGGCGCG CGCGGCTTAC CGCGCGCGCG	420
CTGCGGCGCG AGATTTTTC CGCGCGGCTT AGATTTTTC CGATTTTTC CGCGCGCGCG	480

CGGGGGGGGG GGACCCACCG GTCCCGCCGA TCCCCCCTT GCCGCGGTG CCGCCGCCAT	720
TGCTGCTCT GAAGCCCTTA GCGCGGTTG CGCAGGTTCC GCGCGTGGC CCGTGGCCG	780
CGGGGGGGGG GTTGCCGTAC AGCCACCCCC CGGTGGCGCC GTTGCCGCCA TTGCCGCCAT	840
TGCCGCGCTT GCCGCCATTG CCGCCGTTCC CGCCGCCACC CCGGNTTGG CCGCCGCCG	900
CGCCGGCGGG CGC	913

2) INFORMATION FOR SEQ ID NO:17:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1872 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACCTT GTGTAGAAA ATCTGCGCG CCGGACCTT AAGGCTGGG CAATTTCTGA	60
TAGCTACCC GACACAGGAG GTTACGGGAT GAGCAATTG CCGCGCCGT CACTCAGGTG	120
CTCATGCTT GTGAGCTTC TGGCTGCTT CGGCTGGG GTGGCCACG CCGCGCCCA	180
CGCGCGCCG CCGGCTTGT CGCAGGACG GTTCGCGAC TTCCCGCGG TGCCCTCGA	240
CGCTTCGCG ATGGTCGCG AAGTGGCGG ACAGGTGCT AACATCAACA CCAAACTGG	300
TTAGACAAAG CGCTGGCGG CGCGACCGG CATCTCATC CATCGCAAG GTCTCTGCT	360
TACGACAAAG CAGCTGATC CGGCGCGAC TCACATCAAT CGTTGAGCG TCGGCTCGG	420
CTAAAGCTAC CGCTCGATC TGGTGGGTA CGACCGCAG CAGGATGTC CGGTGCTCA	480
CTGCGCGCT CCGGCTGGG TGGCTGGG CGCATGCTT CGCGCGTTC CGTTGCTGA	540
CTCTCTCTT CGATCGCGA AATCTCTT TACGCGGAA AGGCTCTT CTCTCTCTT	600
CTCTCTCTT CGCTCGCTT AATCTCTT TACGCGGAA AGGCTCTT CTCTCTCTT	660
TAACTTAAAG CGCTGATCG AATCTGATC TCAATCGAG CCGCGTGAT CCGCGCGCG	720
CTCTCTCTT CGCTGATCG AATCTGATC TCAATCGAG CCGCGTGAT CCGCGCGCG	780
CTCTCTCTT CGCTGATCG AATCTGATC TCAATCGAG CCGCGTGAT CCGCGCGCG	840
CTCTCTCTT CGCTGATCG AATCTGATC TCAATCGAG CCGCGTGAT CCGCGCGCG	900

CATCAACTCG GCCACCGCGA TGGCGGACCG GCTTAACCGG CATCATCCCC GTGACGTCAT 1080
 CTCGGTGAAC TGGCAAACCA AGTCGGGGCG CACGCGTACA GGGAACGTGA CATTGGCCGA 1140
 GGGACCCCCG GCGTGATTG TCGCGGATAC CACCCGCGCG CCGGCCAATT GGATTGGCGC 1200
 CAGCCGTGAT TGGCGGTGA GCGCCCGAGT TCGGTCTCCG GTGCGCGTGG CATTGTGGAA 1260
 GCAATGAACG AGGLAGAACA CAGCCTTGAG CACCCTCCCC TGCAGGGCAG TTACSTCGAA 1320
 GCGCGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCGCGCGC CCGCCCGCC 1380
 GATCCGACCT GGTTTAAGCA CGCCGTCTTC TACGAGGTGC TGSTCCGGGC GTTCTTCGAC 1440
 GCGAGCGCGG ACGGTTCCGN CGATCTCCCT GGAATCATCG ATCGGCTCGA CTACCTGCAG 1500
 TGGCTTGGCA TGAATTCAT GTGTTGCGCG GGTTCCTACG ACTCACCCTT GCGCGACCGC 1560
 GGTACGACA TTGCGGACTT CTACAGGTG GTGCGCGAAT TCGGCACCGT CGACGATTTC 1620
 GTGCGCGTGG TCGACACCGC TCACCGGCGA GGTATCGGCA TCATCACCGA CCGGTGATG 1680
 AATCACACCT CGGAGTCGCA CCCCTGGTTT CAGGAGTCCC GCGCGGACCG AGACGGACCG 1740
 TACGGTGAAT ATTACGTGTG GAGCGACACG AGCGAGCGCT ACACCGACCG CCGGATCATC 1800
 TTGCTGACCA CGGAAGAGTC GAACTGGTCA TTGATGCTG TCGCGCGACA GTTCTACTG 1860
 CGACCGATTC TT 1920

2. INFORMATION FOR SEQ ID NO.13

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 1482 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO.13

ATGTCGAA AGCTGATTCG GAGGAACGCG GTGTTGCGCT GAGGCGAAT AGCTGCGAAT
 ATGCTGCT AGCGGAGATC AGCGAGTCCG TCGATCGGAT AAAAGGCTTC AGCAGGCTTC
 AGTACGCGCT CGGAACAACG TGGAAAGTCC ACAGGTTGCT GGGTATTACG AGTCCCGATC
 TGAATGTCG CGCCCAATCG CTGCTGCGAA AGCGGCTATC CATTTACAAC GACGAGCAGG
 GTGCTGCTTC TCGGATACAA TCGATACAA TCGATACAA AGTCTTCA GACTTCAGCA
 TCGGATACAA TCGATACAA TCGATACAA AGTCTTCA GACTTCAGCA

GCGCCCAAGAG TGCAAGGCCG GCGACCGTGT GGATTGCCCA GGACGGCTCG CACCACCTCG 540
 TCCGAGCGAG CATCGACCTC GGATCCGGGT CGATTGAGCT CACSCAGTCG AAATGGAACG 500
 AACCGGTCAA CGTCGACTAG GCCGAAGTTG CGTCGACCGG TTGCTCGAAA CGCCCTTGTG 660
 AACGGTGTCA ACGGCACCCG AAAACTGACC CCCTGACGGC ATCTGAAAAT TGACCCCTTA 720
 GACCGGGCGG TTGGTGGTTA TTCTTCGGTG GTTCCGGCTG GTGGGACGCG GCCGAGCTCG 780
 CGGTCTTTGA GCGCGTAGCT GTCCGCTTTC AGGGCGACGA CTTGAGCATG GTGGACGAGG 840
 CGGTGATCA TGGCGGCAGC AACGACGTCG TCGCCGCCGA AAACCTGCGC CCACCGGCGC 900
 AAGGCGTTAT TGGACGTGAC GATCAAGCTG GCGCGCTCAT ACCGGGAGGA CACCAGCTCG 960
 AAGAAGAGGT TGGCGGCTC GGGCTCAAA GGAATGTAAC CGACTTCGTC AACGACCAGG 1020
 AGCGGATAGC GGGCAAACCG GGTGAGTTTC GCGTAGATGC GCGCGGCGTG GTGAGGCTCG 1080
 GCGAACCCTG CTACCCATTC GCGCGCGGTG GCGAACAGCA CCGGATGACC GGGCTGACAC 1140
 GCGCGTATCG CCAGGCGGAC CGCAAGATGA GTCTTCGCGG TCCGAGGCGG GGGCCAAAAA 1200
 CACGACGTTA TCGCGGCGCG TGATGAAATC CAGGGTGCGC AGATGTGCGA TGGTGTGCGG 1260
 TTTCAGGCCA CGAGCATGCT CAAAGTCGAA GTCTTCGAA CACTTCCGAA CCGGGAAGCG 1320
 GCGCGGCGCG ATGCGGCGCT CAGGACCATG CGACTCGCGG GCTGACACTT CCGGCTGCG 1380
 GCGCGGCGCG AGGATATGCT GGTGCTGCGA GTCTGCGCG TCGCGCGCAT CCGCGAGCG 1440
 TGACGCTGAC TCGCGGCGCG TCGGAGCTTT GAATGCTTTT TT 1480

2. INFORMATION FOR SEQ ID NO.19

a. SEQUENCE CHARACTERISTICS

- A. LENGTH: 1480 base pairs
- B. TYPE: Unlabeled DNA
- C. STRANDEDNESS: Single
- D. TOPOLOGY: Linear

b. SEQUENCE DESCRIPTION: SEQ ID NO.19

GAATTCGCTA CAGCGGCGCG ATAGCTTCTT GCGCGGCGCG CAGCAATATG CTGAGGCTT 80
 GTCTCTGCG GCGCGGCGCG GCGCGGCGCG GCGCGGCGCG CAGCAATATG CAGCGGCGCG 120
 GTCTCTGCG GCGCGGCGCG GCGCGGCGCG GCGCGGCGCG CAGCAATATG CAGCGGCGCG 160

GGAGAACTTC GATCCCGAGG GCGTCTCTGGG GGGTATCTAC CGTATCACC CGGCCACCGA 360
 GCAACGCACC AACAAGGNGC AGATCCTGCC CTCGCGGGTA GCGATGCCCG CGGCGCTGCC 420
 GGCAGCACAG ATGCTGGCCG CCGAGTGGCA TGTCCGCCCC GACGTGTGGT CCGTGACCAG 480
 TTGGGGCGGAG CTAAACCGCG ACGGGGTGGT CATCGACACC GAGAAGCTCC GCCACCCCGA 540
 TCGGCCCGCG GCGGTGCCCT ACGTGACGAG AGCGCTGGAG AATGCTCGGG GCCCGGTGAT 600
 CCGGCTGTCT GACTGGATGC GCGCGGTCCC CGAGCAGATC CGACCGTGGG TGCCGGGCAC 660
 ATACCTCACC TTGGGCACCG ACGGGTTCGG TTTTTCGGAC ACTCGGCCCC CCGGTCTCTG 720
 TTACTTCAAC ACCGACGCGG AATCCGAGT TGGTCCGGGT TTTGGGAGGG GTTGGCCGGG 780
 TCGACGGGTG AATATCGACC CATTCGGTGG CCGTCTCTGG CCGCGCGCCC AATTACGCGG 840
 AATCGACGAA CCGTGGGGGT TCGCGCGCAN TAAGTT 876

2. INFORMATION FOR SEQ ID NO:20:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCCCGCCGG CTCGACGAA TTCCGCACCA GAGACAAAT TCGACGCTT AATCGACGAA 60
 GAGATTGATA ACCAATTGAC AGCCGCACAA GAATATGTTC CGATCCCGGT TTATTTCGAC 120
 ACCBAAGACC TCGTCGAGTT GCGCAAGCAT TTTTACAGGT AAGCGGTGCA GGACGAAAC 180
 CATGCAATGA TCGTCTGCA ACGCTCTCTC GACCGCCACT TTCTGTGCA AATCCCGCC 240
 CTAGACACCG TCGGAAACCA TTTCGACAGA TCGCGCAGG CATTCGCGGT GCGCTCGAT 300
 TCGGAAACCA TCGTCTGCA ACGCTCTCTC GACCGCCACT TTCTGTGCA AATCCCGCC 360
 CATTCGCGGT TCGTCTGCA ACGCTCTCTC GACCGCCACT TTCTGTGCA AATCCCGCC 420
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 480
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 540
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 600
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 660
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 720
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 780
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 840
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 900
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 960
 TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT TTATTTCGAC CGATCCCGGT 1020

GCGGTTGGCC CGACCCCGCT GCGCGCACTG CTGGTCAGGT ATCGGGGGGT GTTGGCGAGC 340
 AACAACTGCG GCAGGAGCGG TGGAGCCCGC CGGATCCGCA GACCGGGGGG CGGAAAACGA 300
 CATCAACACC GCACGGGATC GATCTCGGGA GCGGGGTGCG GGAATACCGA ACCGGGTGTAG 960
 GAGCGCCAGC AGTTGTTTTT CCACCAGCGA AGCGTTTTCG GGTATCGGN GGCNNTTAAG 1020
 T 1021

(2) INFORMATION FOR SEQ ID NO:21:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 321 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTTCCCGACG AACCGAAGAA CACAACCATC AAGATGCTGA AATGATCGT CCGAGCTCTG 60
 ACCGCGCGCG CTGCAATCGG CGCCGCTGCG CCGGCTGTGA GTTCGATCAT GCGTGGCGGN 120
 CCGGCTGTAT ACCAGATGCA CGCGGCTGTC TTGCGCGCGC CACTGCGGTT GGACCGCGNA 180
 TCGGCGCGTC ANGTCCCGAC CGCGCGCGCA TCGACCAGNC TGCTAACAG CTCGCGCGAT 240
 CCGAAGCTGT CGTTTGNAGG CAAGGCGNAGT CTCGTCGAGG GNGGNATCGG NGGNANCGAG 300
 GGNANAGATC NGGNANAGCA A 321

INFORMATION FOR SEQ ID NO:22:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 173 base pairs
 - B. TYPE: nucleic acid
 - C. STRANDEDNESS: single
 - D. TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTTATGCTAT TCGGTTGCG CAGCGGTTCT GGNANAGAT GTTAAAGCG GTGATCGAAT 40
 CATTAAGCG CGCGGAGAG CTGACTCGG ATATTGCGG CGCGTTGAG GTTCAGGCT 100
 CTCGCTGCT GNACTGCGAA CCGCTGAAGG AGCGTTTGA GACCGGATC AAGCGGATG 160
 ATCGATGAG CGGATGCTT CGCGTTCAG GTAAATTA ATCGCGGAG CCGAAGATG 220

CTTACCATCG CCG

373

2' INFORMATION FOR SEQ ID NO:23:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 352 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTGACSCCGT CATGGGATTC CTGGGCGGGG CCGGTCCGCT GCCGGTGGTG GATCAGCAAC 50
 TGGTTAGCCG GGTGCCCCAA GGCTGGTGGT TTGCTCAGGC AGCCGCTGTG CCGGTGGTGT 100
 TCTTGAAGGG CTGGTAGCCG TTCCCGGATT TAGCCGAGAT CAGGCGGGGC GAATCGGTGC 150
 TGATCCATCG CGGTACCCCG GGTGTGGGCA TGGCGGCTGT GCAGCTGGCT GCGCAGTGGG 200
 CGCTGGAGGT TTTCTCACC GCCAGCCGTG GNAAGTGGGA CACGCTCCGC GCCATNGNGT 250
 TTGACGACGA NCCATATCGG NGATTCCGNC ACATNCGAAG TTCCGANGGA GA 352

2' INFORMATION FOR SEQ ID NO:24:

- 1' SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 425 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

x1 SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAAATCCCGG TTGATTGGGT TCGACGAGCG TGTGGCGATA ATCGACGAA TGAATCAGCC 50
 AGGCTTCGCG CGGCTCATCG CTGACAGCGA TTAATCAGCA AGTTCTCTCG TATATCCGAC 100
 CTAGCTGCGA TTGCTTCGCG AGATCGGTTT TGTAGCTCA TCGATCTAA TGGTTCCGCT 150
 TGGCAAGCT CATGCTGCGG TGTGCTATTA TCGGAGCGG TGTGGCGGCT TTGGGCTCG 200
 TGTGCACTA CGCAGCGGAA AGTTCTCTCG TGTGCACTA TGAATCTG TGGGCGAGC 250
 TTGTCGAGCG CGCATCGCGG TCGAAGTGGT ATCGCTACAC CTGCCATGAT GATTCGAGC 300
 TCGACAGCGA CGGCTCGGAC CAGAGTCTCG ATAGCGCGCG AGCCATCTTT GAAAGTCCCG 350
 GATTCGAGCA TCGCTCTCTG GGTGCTGCGG TGTGCGCGG ATATAGCTGT 400
 TGTGCTGCGG TGTGCTGCGG TGTGCTGCGG TGTGCTGCGG TGTGCTGCGG 450

GGCGACAGCG CCGCACCAT CGACATCGAC AAGGTTGTTA CCGGCACACC CGTTCGGCGG 720
ATCGTG 726

(2) INFORMATION FOR SEQ ID NO:25:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCGACGACG ACCAAGCTCG GCGCCACCAC GCGCTATGCG TTGATGCAGG CGACCGGGAT 60
ATCGCGCGAC CATATCGAAG CATCGTGGGT GCGCACTGAG CGACCTTTTG ACCGCGCGGG 120
TTGCGCGATG GCGCGCGGT GAAGTCATTG CCGCGGGGCT TGTGCACCTG ATGAACCCGA 180
ATAGGGAACA ATAGGGGGGT GATTTGCCAG TTCAATGTCG GGTATGGGTG GAAATCCAAT 240
GCGCGGGCAT GCTCGGCGCC GACCAAGGCTC GCGCAGGCGG GCCAGCCCGA ATCTGGAGGG 300
AGCACTCAAT GCGCGCGATG AAGCCCCCGA CCGCGCGACGG TCCTTTGGAA GCAACTAAGG 360
ACCGCGCGCG CATTTGATG CGACTACCAC TTGAGGCTCG CCGTCGCGTG GTCGTCGAGC 420
TGACACCCGA CGAAGCGCG CGACTGGGTG ACGAACGAA AGGCCTTACT AGCTAAGACC 480
AGCGCAAGG CGAATGGTGG KCTTAAGCG CACACCTTCC GGTAGATGTC AGTGTCTTC 540
TGGCGATCT ATGCGCAGGA GAAGTCTTGG ATACAGCGCT 600

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X2 SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAATGAGGG GCGCGCGTTT TCGCGGGGCG GCGCGGTTG GCGCGAAGG GCGCGCGCG 60
GCGCGCGCG GCGCGCGCG TGTGCGCGCG GCGCGCGCG GCGCGCGCG GCGCGCGCG 120
GCGCGCGCG GCGCGCGCG GCGCGCGCG GCGCGCGCG GCGCGCGCG GCGCGCGCG 180

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTCGACA CGCTCGAGGC GTTCACGATC	60
CAGCGCACAC CCGACGGCGT GACCATCGGC GATGCGGCC CGTTCGCGGA GGCGGCTGCC	120
AAGGCGATGG GAATCGACAA GCTGCGGGTA ATTGATACCG GAATGGACCC CGTCGTCGCT	180
GAACGCGAAC AGTGGGACGA CCGCAACAAC ACCTTGGCGT TGGCGCCCGG TGTCGTTGTC	240
GCCTACGAGC GCAACGTACA GACCAACGCC CG	272

2. INFORMATION FOR SEQ ID NO:28:

1. SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCCGGTG GTTCTCGGAC TATCTGCCCA CGGTGACGCA GCGCGAGCTG CCGAGGCTGA	60
ACCGCATCGA GCAGACGGAT CCGTTCGCGG GTTCATGCG GTACCTGGCC GGTATCAGCG	120
CCAGCGAGCT GAACCTGGCT GAAGCGGGCT GGTTCATGCG GTTCGACGCG GCGACGATCC	180
TTTTCGATCT CCGCTGCTT GAGAGCTGT ATTCTGTACA TCGCTTCTT CCGTCTCTGC	240
GAATCTGAC CCGAAGATG AAGAAGGCTT TAAAGATGCA CTCTCTGAG ATGCTCTTCC	300
CTCTCTGCTT CCGCGCG	317

3. INFORMATION FOR SEQ ID NO:29:

1. SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:29:

ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	60
ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	120
ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	180
ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	240
ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	300
ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG ATGATGATG	317

2. INFORMATION FOR SEQ ID NO:30:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 308 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

GATCCCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAAGCGGGT      60
CGGCGTTGAC GAGGCGAAGA CACGCGTGTG CGAGCTGCTG CGGCTCGTCT ACGGCGGGCA      120
TAGGTTGAGA TTGCGCGCGG CGGCGAGCGG GTAGCAAAGC TTGTGCGGCT GCATCGTCAT      180
GAGACTCGGC GATTAGGCAT TGACCATGGC GTGTACCGCG TCGCCGACGA TTTGACGCT      240
CGGTTGTGAG ACGACGTGCT CGAACGCTTT CACCGGTGAA GCGCTACCTC ATCGACACCC      300
ACGTTTGG

```

2. INFORMATION FOR SEQ ID NO:31:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 267 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```

AGAGGAGGA GCAACTGAGT TGAATGATGG TCGGAGGAGG GATTGAGGAC GGAGAGAGTC      60
TGGCGGAGAG TCGCGCGCGG TAACTGATGA TAGTGAGCGG CGGTAGAGGG GTGCGCGGAT      120
AGAGGAGGA TATTGTGCTG TCGGCTGCGG CGGTAAAGAG CGGTAAAAGA ATGTGAGCGG      180
AGAGGAGGA GATGAGAGT TATGAGAGA TGAAGATGCT GGGAGCTG TCGGAGGAGG      240
AGAGGAGGA GATGAGGAGG TGTGTGAG

```

2. INFORMATION FOR SEQ ID NO:32:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1533 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

GCGAGACCAT GCGCGCGCTG GACTGGTTCC AAGTACAGTC AATTCGAGGC CACCTGGTCC 180
 ACAGAGCGGT GCGGCACTTC CAGGTGACTA TGAAAGTCCG CTTCGGCTGG AGGATTCTCTG 240
 AACCTTCAAG GCGGCGCGAT AACTGAGGTG CATCATTAAG CGACTTTTCC AGAACATCCT 300
 GACGCGCTCG AAACGCGGTT CAGCCGACGG TGGCTCCGCC GAGGCGCTGC CTCACAAATC 360
 CCTGCGAATA TCGTGGGGG GCGCCTACAA GGAAGTCGGT GCTGAATTCC TCGGGTATCT 420
 GGTGAGCTG TGTGGGCTGC AGCCGGAAGA AGCGGTGCTC GAGGTGGGCT GCGGCTCGGG 480
 GCGGATGGCG TTGCGGCTCA CCGGTATCTT GAACAGCGAG GGACGCTACG CCGGCTTCGA 540
 CATCTGCGAG AAAGGCTATC CGTGCTGCTA GGAGCACATC ACCTCGGCGC ACCCCAACTT 600
 CCACTTCGAG CTCCTCGACA TCTAATACTC CTTGTACAA CCGAAAGGGA AATACGCTC 660
 ACTAGACTTT GCGTTTCCAT ATCGGATGC CTCGTTGAT GTGGTCTTTC TTACCTGGGT 720
 GTTCAGGAC ATGTTTCCCG CGGAGGTGGA GCAATATCT GACGAGATCT CCGCGCTGCT 780
 GAAGGCGCGC GGAGGATGCC TGTGACGTA CTTCTTGCTC AATGACGAGT CGTTAGGCCA 840
 CATCGCGGAA GGAAAGAGTG CGCACAACCT CCAGCATGAG GGACCGGGTT ATCGGACAA 900
 CCACAAGAA GCGCGCGAAG AAGCAATCG CTTGCGCGAG ACCTTCGTCA GGGATGTCTA 960
 TCGCAAGTTC GCGCTCGCG TCGAAGAAC ATTGCACTAC GGTTCATGGA GTGGCGGGGA 1020
 ACCAGGCTTA AGCTTCAGG ACATGCTCAT GCGACGAA ACCTCGAGCT AGGTTCGGAT 1080
 TCGGGAAGCA TCGGACACT GTGGCGGGA GCGCGCTCC CGCGAGGCTG ATTAGCGCGG 1140
 CAGATTAGCG GCGCGCGGCT CCGGCTGCTC ACTAGGCGCG CGCGAATGCG CTCACCGGCT 1200
 GTAAACGAG CTTGCGCGCG TGGGCGGCG CCGCGCGGAT GAGGTGGTAG ATCCCGACAA 1260
 GAGTATCTC ATGGTCTAT AGCAAGCTC AAGCAAGCTC GTTTCGATC ATCGGCAAGT 1320
 GAGGAGCTC TCGGCTCTT GTCAAGGGA TGAAGGCTC GAGGCTGGA TCGGAGAGT 1380
 GCGGATGAT GTTGGGATC GCGATAGCTT GATAGGAGG ATGAAATTTT GAGGCGGCGA 1440
 GAGATTTTC ATCGGAGAA ACTTGGGCTC GGTTCGGGCT GAGGCGGCTC AGCAGCTTCC 1500
 TGAAGAAAG GAGATTTTC TCGATTTTC GTGTTGCG

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

CTGCAGGGTG GCGTGGATGA GCGTCACCGC GGGGACGGCC GAGCTGACCG CCGCCGAGST      60
CCGGGTTGCT GCGGUGGCGT ACCGACCGGC GTATGGCGTG ACCTGCCCCC CGCCGGTGAT      120
CGCCGAGAAC CGTGCTGAAC TGATGATTCT GATACCGACC AACCTCTTGG GGC AAAACAC      180
CCCGCGGATC GCGGTCAACG AGGCCGAATA CGGCGAGATG TGGGCCCAAG ACGCCGCGGC      240
GATGTTTTGC TACGCGCGCG CGACGCGGAC GCGCAGCGCG ACCTTGCTGG TTTTGGAGGA      300
CGCCCGGGAG ATGACCAAGC GGGTGGGCT GTTGGAGCAG GCGCGCGCGG TCGAGGAGGC      360
TTCCGACACC GCGCGGGCGA ACCAGTTGAT GAAGAATGTG CCGCAGGCGG TGAAGAGTT      420
CGCCAGGCGC ACGGAGGGA GACGCGCTG TTCAAGGTG GGTGGGCTGT GGAAGACGT      480
CTCGCGCGAT CGGTGGCGGA TCAGCAACAT GTTGTGATG GCGAACAAGT ACATGTGAT      540
GACCAACTCG GGTGTGTGGA TGACCAACAT GTTGAGCTCG ATGTTGAAGG GCTTTGCTCC      600
CGCGCGGCGC GCGCAGGCGG TCAGAACCGT GCGCGAAAAC GGGGTCCGCG TATGAGGTC      660
GCTGGGACCG TCGGTGGCTT GTTGGGCTGT TGGGGTGGG GTGGCGCGCA ACTTGGGTCG      720
CGCGCGCTCG GTAGGTATG GTACCGGGA TGGCGAAAA TATGCANAGT CTGCTGCGCG      780
GAAGGTGCTT CGGGGTAAAG GTTACCGCG GTTTGTGGA TCGGTGAAG GTGTCAGG      840
GAAGAGTTTA C

```

1. INFORMATION FOR SEQ ID NO:34:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 294 base pairs
- B. TYPE: nucleotide
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

ATTGATCGG CGGAAAATT GACGAGATT GTTGTGCTT GATAAGGAA TGAATCGAA      60
GTAGATTTAT TCGGTGAGG GCGCGAGTA ATGCTGCGA GAGAGGAA GTTATGCTG      120
GCTGAGCTG GGTAGATTG GTATAGGAA GAAAGCTT GATATTTT GATAGAGT

```

(x) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 1227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```

GATCCTGACC GAAGCGGCCG CCGCCAAGGC GAAGTCCCTG TTGGACCAGG AGGGACGGGA      60
CGATCTGGCG CTGCGGATCG CGGTTTCAGCC GCGGGGGGTGC GCTGGATTGC GCTATAACCT      120
TTTCTTGGAC GACCGGACCG TGGATGGTGA CCAAAACCGC GAGTTCCGTG GTGTGAGGTT      180
GATCTGTGAC CGGATGAGCG CGCGGTATGT TGAAGGCGCG TCGATCGATT TCGTCGACAC      240
TATTGAGGAG GAAGGTTGAG GATCGACAAT TCAAAACCGA CCGGCTCTCT GCGCTCCCGG      300
GATTGCTTCA ACTGATAAAA CGGTAGTAGG ACCCGCGCGT GCGCAACAGG TACGAGTACA      360
CCAAAGAGTG ACCGCGCTGG AAAAGCAACT GAGCGATGCG TTGCACTTGA CCGGCTGGCG      420
GGCGCGCGCG GCGAGGTGTC ACCTGCATGG TGAACAGCAC CTGGGCTTGA TATTGCGACC      480
AGTACAGCAT TTTGTGGATC GAGGTCACCT CGAGCTGGGA GAACTGCTTG CGGAACGCGT      540
CGCTGCTGAG GTTGGCCAAG CGGTGATCGG AGCGCTTCTC GCGCAGCGCG TCGTGGATAC      600
CGCAGAGCGG ATTGCGAAGG ATGCTGTGCA GATCGCGCTT CTGAGTGGCG TTGAGGATAC      660
CTGGAATGCG GCTTTTGGCG CGTCCCTCGG AGAATCTGCG TCGCTGTCTG GCTGCTTTGG      720
TGGGAGCGCG GTATATGATG CGCGCGCTCA TACCGGACAG CAGCGCGAGG GCTACGACAA      780
TGGCGATCAG CAGCGCGCTG TCGCTGCGCT TCGGCTAGGA CAGCTGCGCG GCGACGCGCG      840
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      900
CGCGCGCGAG TCGCTGCGCT GATCGAGTGT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      960
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      1020
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      1080
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      1140
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      1200
GATATCGCGG CGCGCGCGAG TCGCTGCGCT TTGCGCGTGC CGCGGCGAAG CGCGCTTGGG      1260

```

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

GCGGTGTCCG CGCATCCGGC GGGTGGTTGA ACGGCAACGG CGGGGCCGGC GGGGCCGGCG      60
GGACCGGGCG TAACGGTGGT GCCGJCGGCA ACGCCTGGTT GTTCGGGGCC GCGCGTCCG      120
GCGGNGCCCG CACCAATGGT GGNGTCGGCG GTCGGGCGG ATTTGTCTAC GGCAACGGCG      180
1                                           191
  
```

2 INFORMATION FOR SEQ ID NO:37:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 290 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```

GCGGTGTCCG CGCATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCCG GCGCGGGCG      60
GGACCGGGCG TAACGGTGGT GCCGJCGGCA ACGCCTGGTT GTTCGGGGCC GCGCGTCCG      120
GCGGNGCCCG CACCAATGGT GGNGTCGGCG GTCGGGCGG ATTTGTCTAC GGCAACGGCG      180
GGGTGATCCG CGCATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCCG GCGCGGGCG      240
GGACCGGGCG TAACGGTGGT GCCGJCGGCA ACGCCTGGTT GTTCGGGGCC GCGCGTCCG      300
  
```

3 INFORMATION FOR SEQ ID NO:38:

SEQUENCE CHARACTERISTICS
 A. LENGTH: 14 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

4. SEQUENCE DESCRIPTION: SEQ ID NO:38:

```

ATTGCAATGG CATGGGGGGT CTCAGTGGAA GCAT      14
  
```

5 INFORMATION FOR SEQ ID NO:39:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 155 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

GATGGCTGCT CGTCCCCCCC TTGCCCCCGA CCCCACCGGT CCCACCGTTA CCGAACAAGC 60
 TGGCGTGGTC GCCAGCACCG CCGGCACCGC CGACGCCGGA GTCGAACAAT GGCACCGTCG 120
 TATCCCCACC ATTGCCCCCG GNCCCACCGG CACCG 155

(2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATCCCCTTCA CCGGGCCCCG GGCACCGCGG AGCCCCGNGG CCCCCCGGGG TCG 63

(2) INFORMATION FOR SEQ ID NO:41:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATTACACCG CGTTCAGAC CGTCCCCCGG CCGCACCGCG GACCAGCGCG CGTAACGGCG 60
 CATTTCCTTC CAACTTCGCG AACCGCACCG CGTTCGNGG CCCCCCGCGG CCGCGCGGCA 120
 CCGCGCGCAA TC 132

(2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATCTCCCGT CGTAACCGCG CGGTACCGCG CCAAGCGCGG NAACCGCGCG CCGTAGCGCA 60
 CAGCGCAAGA ATCTTCGCGG CGTTCGCAAT CTAAGAATCG CGACACCGCG CGTAACCGCG 120
 CAGTTCGCGG CA 132

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO 43:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCCGCGGG TTTCGCCACC	60
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCC CGATGCCGGC	120
ATCAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCACT TTAGCGACGA TAATGGCTAT	180
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG	240
AGATTTTGAA CAGGGCCAAC GAGGTGGAGG CCCCAGTGGC GGACCCACCG ACTGATGTCC	300
GCATCAGACC GTCCGAATC ACCGNGNTA AAAACGCCCG CCAACAGNTG GTNTGTCTCG	360
CCGACAACAT CGCGCAATAC CTGCGCGCGG GTGCCAAGA CGCGCAGCGT CTGGCGACCT	420
CGGTGCGCAA CGCGGCCAAG GNGTATGGCG AGGTTGATGA CGAGGCTGCG ACCGCGCTGG	480
ACAACGACCG CGAAGGAAC GTGCAGGCAC AATCGGCCCG GCGCGTCGGA GGGGACAGTT	540
CGGCGCAACT AACCGATACG CGGAGGCTCG CCACGCGCGG TGAACCCAAC TTCATGGATC	600
TGAAAGAAAG CGCAAGGAAG CTCGAACCG CGCACCAAGG CGCATCGCTC CGCGACTGNG	660
CGGATGGCTG GAACASTTC AGGCTGACCT TCGAAGCGCA CG	702

(x2) INFORMATION FOR SEQ ID NO 44:

SEQUENCE CHARACTERISTICS:
 A. LENGTH: 398 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(x3) SEQUENCE DESCRIPTION: SEQ ID NO 44:

AAATGAGG AGCTGTCCCG TAACTGAT TGAAGATG ATGTTCTCGT TCGCTCTGGA	
ATGTTCTCGT TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG	12
TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG	18
TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG	24
TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG TCGCTCTGCG	30

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```

CGGCACGAGG ATCGAATCGC GTCGCCGGGA GCACAGCGTC GCACTGCACC AGTGGAGGAG      60
CCATGACCTA CTCGCCGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG      120
GAGGCGTCAC ACCCTCGTTC GCCCAGCCCG ATGAGGGTGC GAGCAAGCTA CCGATGTACC      180
TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTCCCTACTT CGCCAGCTTC GGCCTAATGT      240
TCACCTTCAG TACCGAACTC GGGGGGGGTG ATGGCCCACT GTCCGGTGAC ACTGGGCTGC      300
CGCTCCCGCT CGCTATCTTC GCTGCCCTGC TTCCCGCGGT GGTTCCTGGT CCAAGGCCA      360
AGAGGCATGT GACGGTAGTT GCGGTGCTCG GGGTACTCGG CGTATTTCTG ATGGTCTCGG      420
CGAGCTTTAA CAAGCCCGAGC GCTATTCTGA CGCGTTGGGC ATTGTGGCTT CTGTTGGCTT      480
TCATCTGTGT CCAGGCGGTT GCGGCAGTCC TGGCGCTGTT GGTGGAGACC GCGGCTATCA      540
CGCGCCCGGC CGCGCGGCCC AAGTTGAGC CSTATGGACA GTACGGGCGG TACCGGCACT      600
ACCGGCAGTA CGGGGTGCAG CGGGGTGGGT ACTACGGTCA GCAGGGTGCT CAGCAGCCCG      660
TGGCACTGCA GTCCGCCCGC CGGCAGCAGT CTCGCAGCC TCCCGGATAT GGGTGCAGT      720
ACCGCCGCTA TTCCTCCAGT CGAGCCCAAT CGGCAGTGG ATACACTGCT CAGGCCCGGG      780
CCCAJCCGCG GCGGCAGTCC GGTGCGAAC AATCGCACCA TGGCCCATCC AAGCCAGCTA      840
AGGCTTTTCC GAGCTTCAGC GCACCACCAC CGGTCACTTC TGGACGCGT TCGAGGCTG      900
ATCGGCTTCC AGTCACTAT TCAAACCCCA CGGGGCGGCA CGAGTCTTCC TCGCCCGGGG      960
AGGCTTGGGT CTAAGCCCGC GTTCCCGGCT ATGCTCGGCT GTTCTCTGCA AGATGAAGA      1020
AGGCTTGGGT CTAAGCCCGC GTTCCCGGCT ATGCTCGGCT GTTCTCTGCA AGATGAAGA      1080

```

(x1) INFORMATION FOR SEQ ID NO:46:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 127 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO:46:

AGTGGCGCGG CGCGGCGGGG ACGGCGGCGG AGGCGGCGGT GGTGCGCTTC CAAGAAGCAG 180
 CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTCGACGAA TATTCGTCAG GCCGGCCTCC 240
 AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCGCTGTC CTCGCAAATG GGCTTCTGAC 300
 CCGCTAATAC GAAAAGAAAC GGAGCAA 327

(2) INFORMATION FOR SEQ ID NO:47:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 170 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1: SEQUENCE DESCRIPTION: SEQ ID NO:47:

CGGTGCGGAT GATGGCGCTG TCGAACSTGA CGGATTCTGT ACCGCGCTCG TTCAGATCAA 90
 CCAACAACCT GTTGGCGCTG CCAAATGTGG CGNACCCGTG GATCTCGGTG ATCTTCTTCT 120
 TCTTCATCAG GAAGTGACA CCGGCCACCG TCGCTCGGN TACCTTTTCGG 170

(2) INFORMATION FOR SEQ ID NO:48:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1: SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCGCGCGG CACGGGGGGT GCGGCGCGCA CGACCGCTCG CGGTGGCGGT AACGGCGGGG 60
 CGCGGGGTGG CGCGGGAACC GTGGGTTGG TCTTGGGCAA CGCGGTGGT GCGGGGACG 120
 CGCGGT 127

(2) INFORMATION FOR SEQ ID NO:49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1: SEQUENCE DESCRIPTION: SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 149 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GATCAGGGCT GCGCGGCTCC GCGCAGAAGG GCGGTAACGG AGGAGCTGCC GGATTGTTTG 60
 GCAACGGCGG GCGCGGNGGT GCGCGGCGCT CCAACCAAGC CGGTAACGGC GNGCGCGGCG 120
 GAAACGGTGG TGCGGGTGGG CTGATCTGG 149

(i) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC TGAGTGATG AGATGCTCGG GACCTCGCCC GACGGTGTCTG 60
 AGCGCGNAAT CGAGGGCGGT CTGGCGCGAG CTGCGCAGAC CATCGCGCGG CTGGACTGGT 120
 TCGAAGTACA CTCGAATTCA GCGACCTCG TCGACCGAGC GGTGCGCGAC TTCCAGGTGA 180
 CTATGAAAGT CGGCTTCGCG CTGGAGGATT CTGGAAGCTT CAGCGCGCGC GCATAACTGA 240
 CTGCGATCAT TAAGCGACTT TTGCAAAAGA CTGCAAGCGG CTCGAAAGCG GTTCACCGCG 300
 AGGTTGCGTC CGCGGAGGCG CTGCTGCGAA AATGCTCGG ACAATTGCTG GCGCG 355

(i) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 399 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CTGATTAAG AGGAGAGCA CAGCAGCGAG CTGCAAGCGA ATTGATA CTGCAAGCGA 60
 CTGATTAAG AGGAGAGCA CAGCAGCGAG CTGCAAGCGA ATTGATA CTGCAAGCGA 120

```

GCCCCCGCCA ACACGCGGAA TCCCCAGCCG GCGGATCCCA ACGCAGCACC TCCGCGCGCC      300
GACCGGAACG CACCGCGCGC ACCTGTCATT GCGCCAAACG CACCCCAACC TGTCGGGATC      360
GACAACCCCG TTGGAGGATT CAGCTTCGCG CTGCTGCTG GCTGGGTGGA GTCTGACGCC      420
GCCCCACTTC ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCAATTCCC      480
GGACAGCCGC CGCCGGTGGC CAATGACACC CGTATCGTGC TCGGCCCGCT AGACCAAAAG      540
CTTTACGCCA GCGCCGAAGC CACCGACTCC AAGGCGCGCG CCGGGTTGGG CTGGGACATG      600
GGTGAGTTCT ATATGCGCTA CCGCGGCACC CGGATCAACC AGGAAACCGT CTCGCTCGAC      660
GCGAACCGGG TGTCTGGAAG CCGCTGCTAT TACGAAGTCA AGTTCAGCGA TCCGAGTAAG      720
CGGAACCGCC AGATCTGGAC GGGCCTAATC GGTCTGTCGG CCGCGAACCG ACCCGACGCC      780
GCGCCCCCTC AGCGCTGGTT TGTGGTATGG CTGGGGACCG CCAACAACCG GGTGGACAAG      840
GCGCGCGCCA AGGCCTGCGC CGAATCGATC CCGCTTTTGG TCGCCCCCGC GCGCGCGCGC      900
GCACCGGCTC CTGCAGAGCC CGCTCCCGCG CCGCGCCCGG CCGGGGAAGT CGCTCTACC      960
CGGACGACAC CGACACCGCA CCGGACCTTA CCGGGCTGA      999

```

2. INFORMATION FOR SEQ ID NO.33:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 332 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO.33

```

Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1           5           10           15
Asp Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val
16          20          25          30          35          40          45
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
46          50          55          60          65          70          75
His Thr Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
81          85          90          95          100          105          110
Ala Ala Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
111         115         120         125         130         135         140

```

100	105	110
Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser		
115	120	125
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp		
130	135	140
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro		
145	150	155
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg		
165	170	175
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala		
180	185	190
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro		
195	200	205
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val		
210	215	220
Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys		
225	230	235
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn		
245	250	255
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly		
260	265	270
Thr Ala Asn Asn Pro Val Asp Lys Ily Ala Ala Lys Ala Leu Ala Glu		
275	280	285
Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro		
290	295	300
Ala His Ala Ala Thr Ala Pro Val Lys Ala Gly Glu Val Ile Thr Thr		
305	310	315
Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr		
320	325	330

1. INFORMATION FOR SEQ ID NO: 14

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 33 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:

2) INFORMATION FOR SEQ ID NO:55:

001 SEQUENCE DESCRIPTION: SEQ ID NO:55:

2 INFORMATION FOR SEQ ID NO: 56

01 SEQUENCE DESCRIPTION. SEC ID NO: 54.

ALL DAY LONG

1. INFORMATION FOR SEQ ID NO: 57

RECEIVED: 1987 FEBRUARY 24

EXPIRATION: FEB 28, 1974

CONCLUSIONS

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
1 5 10

(x) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Gln Ser Ile Ser Thr Xaa Gln Xaa Ile Val Pro
1 5 10

(xii) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Ala Ala Ala Ala Pro Pro
1 5 10 15

Ala

(xiii) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Ile Gly Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(xiv) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Gln Thr Ser
1           5           10           15

Leu Leu Asn Asn Leu Ala Asp Pro Asp Val Ser Phe Ala Asp
20           25           30

```

(2) INFORMATION FOR SEQ ID NO:63:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:63:

```

Gly Cys Gly Asp Arg Ser Gly Gly Asn Leu Asp Gln Ile Arg Leu Arg
1           5           10           15

Arg Asp Arg Ser Gly Gly Asn Leu
20

```

(2) INFORMATION FOR SEQ ID NO:64:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 187 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:64:

```

Thr Gly Ser Leu Asn Gln Thr His Asn Arg Arg Ala Asn Gln Arg Cys
1           5           10           15

Asn Thr Thr Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala
20           25           30           35

Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val
35           40           45           50           55           60

Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro
65           70           75           80           85           90

Leu Pro Leu Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln
95           100          105          110          115          120

Leu Thr Ser Leu Leu Asp Ser Leu Ala Asp Thr Thr Val Ser Thr Ala
125          130          135          140          145          150

```

```

Ile Ala Asp His Lys Leu Lys Lys Ala Ala Glu His Gly Asp Leu Pro
  115                      120                      125

Leu Ser Phe Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala
  130                      135                      140

Thr Ala Asp Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr
  145                      150                      155                      160

Gln Asn Val Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala
  165                      170                      175

Ser Ala Met Glu Leu Leu Gln Ala Ala Gly Xaa
  180                      185

```

2) INFORMATION FOR SEQ ID NO:65:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 148 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```

Asp Glu Val Thr Val Glu Thr Thr Ser Val Phe Arg Ala Asp Phe Leu
  1           5           10           15

Ser Gln Leu Asp Ala Pro Ala Gln Ala Gly Thr Glu Ser Ala Val Ser
  20           25           30

Gly Val Gln Gly Leu Pro Pro Gly Ser Ala Leu Leu Val Val Lys Arg
  35           40           45

Gly Pro Asn Ala Gly Ser Arg Phe Leu Leu Asp Gln Ala Ile Thr Ser
  50           55           60

Ala Gly Arg His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val
  65           70           75           80

Ser Arg Arg His Ala Val Phe Arg Leu Gln Asn Asn Thr Phe Asn Val
  85           90           95

Val Asp Val Gly Ser Leu Asn Gly Thr Thr Val Asn Arg Gln Pro Val
  100          105          110          115

Asp Ser Ala Val Leu Ala Asn Gly Asp Gln Val Gln Ile His Lys Leu
  120          125          130          135

Arg Leu Val Phe Leu Thr Gly Pro Asn Gln Thr Gly Asp Asn His Ser
  140          145          150          155          160

```

(2) INFORMATION FOR SEQ ID NO:66:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 230 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr
 1 5 10 15

Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
 20 25 30

Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser
 35 40 45

Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
 50 55 60

Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu
 65 70 75 80

Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu
 85 90 95

Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser
 100 105 110

Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp
 115 120 125

Gln Lys Ser Leu Glu Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu
 130 135 140

Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn
 145 150 155 160

Leu Thr Ser Ala Thr Tyr Gln Ser Ala Ile Pro Ser Arg Gly Thr Ile
 165 170 175

Ala Glu Val Leu Asn Val Cys His Asn Ala Gly Gly Thr His Pro Thr
 180 185 190

Thr Thr Thr Lys Ala Phe Asp Trp Asp Gln Ala Tyr Arg Lys Pro Ile
 195 200 205

Thr Thr Asp Thr Ser Thr Gln Ala Asp Thr Asp Pro Leu Pro Val Val

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
1           5           10           15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
20           25           30

Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35           40           45

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
50           55           60

Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
65           70           75           80

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
85           90           95

Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
100          105          110

Ile Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Gln
115          120          125

Gly Pro Pro Ala
130

```

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```

Val Pro Leu Arg Ser Pro Ser Met Ser Pro Ser Lys Tyr Leu Ala Ala
1           5           10

```

Met Ala Arg Val Arg Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa
 50 55 60

Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val
 65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly
 85 90 95

Ser Glu Arg Lys
 100

2. INFORMATION FOR SEQ ID NO:69:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr
 1 5 10 15

Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu
 20 25 30

Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp
 35 40 45

Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Thr Tyr Ala Gly
 50 55 60

Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Val Gly His Leu
 65 70 75 80

Thr Ala Val Leu Ile Val Asn Val Ala Ile Asp Arg Arg Asp Ile Arg
 85 90 95

Arg Arg Leu Ala Arg Arg Ser Thr Thr Val Val Leu Arg Ala Arg
 100 105 110

Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly
 115 120 125

Ala Ala Ala His Leu Gly Thr Thr Gly Val Leu Ala Ala Leu Gly Arg
 130 135 140 145

2 INFORMATION FOR SEQ ID NO:70:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 344 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO:70:

```

Met Lys Phe Val Asn His Ile Glu Pro Val Ala Pro Arg Arg Ala Gly
 1             5             10             15

Gly Ala Val Ala Glu Val Tyr Ala Glu Ala Arg Arg Glu Phe Gly Arg
 20             25             30

Leu Pro Glu Pro Leu Ala Met Leu Ser Pro Asp Glu Gly Leu Leu Thr
 35             40             45

Ala Gly Trp Ala Thr Leu Arg Glu Thr Leu Leu Val Gly Gln Val Pro
 50             55             60

Arg Gly Arg Lys Glu Ala Val Ala Ala Ala Val Ala Ala Ser Leu Arg
 65             70             75             80

Lys Pro Trp Cys Val Asp Ala His Thr Thr Met Leu Tyr Ala Ala Gly
 85             90             95

Gln Thr Asp Thr Ala Ala Ala Ile Leu Ala Gly Thr Ala Pro Ala Ala
100            105            110

His Asp Pro Asn Ala Pro Tyr Val Ala Trp Ala Ala Gly Thr Gly Thr
115            120            125

Pro Ala Gly Pro Pro Ala Pro Phe Gly Pro Asp Val Ala Ala Glu Tyr
130            135            140

Leu Gly Thr Ala Ala Gln Phe His Phe Ile Ala Arg Leu Val Leu Val
145            150            155            160

Leu Leu Asp Gly Thr Ser Ser Val Gly Phe Ser Arg Ala Ala Gln Leu
165            170            175

Met Arg Arg Ala Gly Gly Leu Val Phe Ala Arg Lys Val Arg Ala Glu
181            185            190

His Arg Pro His Arg Ser Thr Arg Arg Leu His Pro Arg Thr Leu Pro
195            200            205

```

Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro
 245 250 255
 Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro
 260 265 270
 Ala Asp Leu His Ala Pro Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala
 275 280 285
 Pro His Gln Val Thr Asp Asp Asp Val Ala Ala Ala Arg Ser Leu Leu
 290 295 300
 Asp Thr Asp Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Phe Thr
 305 310 315 320
 Ala Ala Arg Arg Ile Gly Thr Trp Ile Gly Ala Ala Ala Glu Gly Gln
 325 330 335
 Val Ser Arg Gln Asn Pro Thr Gly
 340

12 INFORMATION FOR SEQ ID NO:71:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 485 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:71:

Asp Asp Pro Asp Met Pro Gly Thr Val Ala Cys Ala Val Ala Asp Ala
 1 5 10
 Leu Gly Arg Gly Ile Ala Pro Val Gln Asp Ile Gln Asp Cys Val Glu
 15 20 25 30
 Ala Arg Leu Gly Gln Ala Gly Leu Ser Asp Val Ala Arg Val Tyr Ile
 35 40 45 50
 Gly Tyr Arg Gln Arg Arg Ala Gln Val Arg Thr Ala Val Ala Leu Leu
 55 60 65
 Gly Val Arg Asp Gln Leu Cys Leu Ser Leu Ala Ala Val Thr Val Leu
 70 75 80
 Arg Gln Arg Tyr Leu Leu His Asp Gln Gln Gly Arg Pro Ala Val Ser
 85 90 95

Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met
 130 135 140
 Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro
 145 150 155 160
 Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala
 165 170 175
 Glu Leu Gln Arg Ala Gly Gly Gly Thr Gly Tyr Ala Phe Ser His Leu
 180 185 190
 Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly
 195 200 205
 Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser
 210 215 220
 Met Gly Gly Arg Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser
 225 230 235 240
 His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser
 245 250 255
 Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu
 260 265 270
 Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr
 275 280 285
 Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Leu Phe Asp Ala Ile
 290 295 300
 Lys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp
 305 310 315 320
 Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala
 325 330 335
 Thr Asn Pro Lys Gly Glu Val Pro Leu Leu Pro Tyr Ile Ser Thr Asn
 340 345 350
 Met Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp
 355 360 365
 Asp Asp Arg Leu Glu Glu Val Ala Gly Val Ala Val Arg Phe Leu Asp
 370 375 380
 Asn Val Ile Asn Val Ser Arg Tyr Pro Phe Ile Glu Leu Ala Ala

Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg
 420 425 430
 Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala
 435 440 445
 Ser Arg Arg Leu Ala Glu Glu Arg Gly Ala Phe Pro Ala Phe Thr Asp
 450 455 460
 Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala Gln Val Thr Ser
 465 470 475 480
 Val Ala Pro Thr Gly
 485

(xii) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 267 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Gly Val Ile Val Leu Asp Leu Gln Pro Arg Gly Pro Leu Pro Thr Glu
 1 5 10 15
 Ile Tyr Trp Arg Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val Val
 20 25 30
 Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala
 35 40 45
 Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His
 50 55 60
 Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu
 65 70 75 80
 Ile Asn Ala Ala Ala Ala Pro Arg Gln Ile Gln Asn Pro Gln Thr Pro
 85 90 95
 Thr Pro Thr Ala Ala Val Gln Glu Pro Pro Val Leu Lys Gln Gly Asp
 100 105 110
 Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro
 115 120 125
 Gln Tyr Tyr Val Gln Asn Gln Pro Lys Phe Thr Met Val Val Thr Asn

Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp
 165 170 175

Cys Ala Pro Ser Asn Glu Thr Leu Val Lys Thr Phe Ser Pro Gly Glu
 180 185 190

Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg
 195 200 205

Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val
 210 215 220

Val Gln Leu Gly Asn Leu Arg Ser Leu Pro Val Pro Phe Ile Leu Asn
 225 230 235 240

Gln Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln
 245 250 255

Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly
 260 265

1. INFORMATION FOR SEQ ID NO:73:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly Val Gln Val
 1 5 10 15

Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Gln Val Val Ala
 20 25 30

Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val Val Val Thr
 35 40 45

Lys Val Asn Asp Ala Ser Ile Asn Ser Ala Ser Ala Leu Val Ala Ala
 50 55 60

Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr Pro Gln Asp
 65 70 75 80

Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly Lys Ala Gln
 85 90 95

Gln

- (A) LENGTH: 364 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala
1           5           10           15

Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser
          20           25           30

Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser
          35           40           45

Thr Ala Gln Glu Asn Ala Met Glu Gln Phe Val Tyr Ala Tyr Val Arg
          50           55           60

Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala
          65           70           75           80

Gly Val Thr Gln Phe Leu Asn Asn Glu Thr Asp Phe Ala Gly Ser Asp
          85           90           95

Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg
          100          105          110

Cys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala
          115          120          125

Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro
          130          135          140

Thr Thr Ala Lys Ile Phe Asn Gly Thr Glu Thr Val Trp Asn Asp Pro
          145          150          155          160

Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Pro Thr Pro Ile
          165          170          175

Ser Val Ile Phe Asn Ser Asp Lys Ser Thr Thr Ser Asp Asn Pro Gln
          180          185          190

Lys Tyr Leu Asp Gly Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser
          195          200          205

Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly
          210          215          220

Thr Ser Ala Leu Leu Gln Thr Thr Arg Glu Leu Ile Thr Thr Asn Gln

```

Ser Ala Gly Pro Asp Pro Val Ala Ile Thr Thr Glu Ser Val Gly Lys
 260 265 270
 Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu
 275 280 285
 Asp Thr Ser Ser Phe Tyr Arg Pro Thr Gln Pro Gly Ser Tyr Pro Ile
 290 295 300
 Val Leu Ala Thr Tyr Glu Ile Val Cys Ser Lys Tyr Pro Asp Ala Thr
 305 310 315 320
 Thr Gly Thr Ala Val Arg Ala Phe Met Gln Ala Ala Ile Gly Pro Gly
 325 330 335
 Gln Glu Gly Leu Asp Gln Tyr Gly Ser Ile Pro Leu Pro Lys Ser Phe
 340 345 350
 Gln Ala Lys Leu Ala Ala Ala Val Asn Ala Ile Ser
 355 360

(1) INFORMATION FOR SEQ ID NO:75:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 369 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Gln Ala Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Gln Asp
 1 5 10 15
 Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val
 20 25 30
 Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro
 35 40 45
 Pro Arg Arg His Pro Ala His His His Arg Arg Arg Val Ala His Ser
 50 55 60
 Gly Gly Arg Arg Arg Pro His Pro His His Val Gln Pro Asp Asp Arg
 65 70 75 80
 Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Pro Ala Gln His Pro
 85 90 95
 Asn Pro His Arg Ala Gln Pro Ala Asp Pro Gly Arg Val Arg Gly Arg

Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val
 130 135 140

Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg
 145 150 155 160

Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly
 165 170 175

Asp Val Leu Arg Ala Gly Leu Arg Val Glu Arg Leu Arg Pro Val Ala
 180 185 190

Ala Val Glu Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val
 195 200 205

Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg
 210 215 220

Ala Gly Pro Gln Gly Arg Leu His Leu Asp Gly Ala Gly Pro Ser Pro
 225 230 235 240

Leu Pro Ala Arg Ala Gly Gln Gln Gln Pro Ser Ser Ala Gly Gly Arg
 245 250 255

Arg Ala Gly Gly Ala Glu Arg Ala Asp Pro Gly Gln Arg Gly Arg His
 260 265 270

His Gln Gly Gly His Asp Pro Gly Arg Gln Gly Ala Gln Arg Gly Thr
 275 280 285

Ala Gly Val Ala His Ala Ala Ala Gly Pro Arg Arg Ala Ala Val Arg
 290 295 300

Asn Arg Pro Arg Arg
 305

2. INFORMATION FOR SEQ ID NO:16:

2.1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 30 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2.2. SEQUENCE DESCRIPTION: SEQ ID NO: 16

Asn Ala Val Thr Cys Leu Asn Gly Pro Thr Gly Arg His Arg His Gly
 1 10 20

Arg Gly Arg Val Arg Ala Ser Gly Thr Arg Ser Ser Asn Arg Thr Cys
 21 30

Ser Pro Leu Glu Arg Arg Phe Thr Cys Cys Ser Pro Ala Val Gly Cys
 50 55 60

Arg Phe Arg Ser Phe Pro Val Arg Arg Leu Ala Leu Gly Ala Arg Thr
 65 70 75 80

Ser Arg Thr Leu Gly Val Arg Arg Thr Leu Ser Gln Trp Asn Leu Ser
 85 90 95

Pro Arg Ala Gln Pro Ser Cys Ala Val Thr Val Glu Ser His Thr His
 100 105 110

Ala Ser Pro Arg Met Ala Lys Leu Ala Arg Val Val Gly Leu Val Gln
 115 120 125

Glu Glu Gln Pro Ser Asp Met Thr Asn His Pro Arg Tyr Ser Pro Pro
 130 135 140

Pro Gln Gln Pro Gly Thr Pro Gly Tyr Ala Gln Gly Gln Gln Gln Thr
 145 150 155 160

Tyr Ser Gln Gln Phe Asp Trp Arg Tyr Pro Pro Ser Pro Pro Gln
 165 170 175

Pro Thr Gln Tyr Arg Gln Pro Tyr Glu Ala Leu Gly Gly Thr Arg Pro
 180 185 190

Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gly Met
 195 200 205

Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr
 210 215 220

Ile Ala Val Val Ser Ala Gly Ile Gly Gly Ala Ala Ala Ser Leu Val
 225 230 235 240

Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala
 245 250 255

Val Ala Ala Thr Ser Thr Ala Ala Asn Met Ser Ser Gly Ser Val
 260 265

Glu Gln Val Ala Ala Leu Val Val Pro Ser Val Thr Met Leu Thr Thr
 270 275 280

Asp Leu Gly Arg Gln Ser Glu Gln Gly Ser Gly Ile Ile Leu Ser Ala
 290 295 300

Val Gly Ser Ile Thr Thr Asn Asn His Val Ile Ala Ala Ala Ala Lys
 305 310 315 320

340 345 350
 Ile Ala Val Val Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser
 355 360 365
 Leu Gly Ser Ser Ser Asp Leu Arg Val Gly Gln Pro Val Leu Ala Ile
 370 375 380
 Gly Ser Pro Leu Gly Leu Glu Gly Thr Val Thr Thr Gly Ile Val Ser
 385 390 395 400
 Ala Leu Asn Arg Pro Val Ser Thr Thr Gly Glu Ala Gly Asn Gln Asn
 405 410 415
 Thr Val Leu Asp Ala Ile Gln Thr Asp Ala Ala Ile Asn Pro Gly Asn
 420 425 430
 Ser Gly Gly Ala Leu Val Asn Met Asn Ala Gln Leu Val Gly Val Asn
 435 440 445
 Ser Ala Ile Ala Thr Leu Gly Ala Asp Ser Ala Asp Ala Gln Ser Gly
 450 455 460
 Ser Ile Gly Leu Gly Phe Ala Ile Pro Val Asp Gln Ala Lys Arg Ile
 465 470 475 480
 Ala Asp Gln Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly
 485 490 495
 Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Gln
 500 505 510
 Val Val Ala Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val
 515 520 525
 Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu
 530 535 540
 Val Ala Ala Val Arg Ser Lys Ala Arg His Ala Thr Val Ala Leu Thr
 545 550 555
 Leu Val Asp Arg Ser Gly Lys Ser Arg Thr Val Gln Val Thr Leu Gln
 560 565 570
 Lys Ala Gly Gln
 580

1. INFORMATION FOR SEQ ID NO 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 580

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```

Met Asn Asp Gly Lys Arg Ala Val Thr Ser Ala Val Leu Val Val Leu
 1             5             10             15

Gly Ala Cys Leu Ala Leu Trp Leu Ser Gly Cys Ser Ser Pro Lys Pro
 20             25             30

Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr Ala Ser Asp Pro
 35             40             45

Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala Thr Lys Gly Leu
 50             55             60

Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys Val Asp Ser Leu
 65             70             75             80

Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala
 85             90             95

Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg
100             105             110

Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn
115             120             125

Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
130             135             140

Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145             150             155             160

Gly Thr His Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr
165             170             175

Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala
180             185             190

Arg Pro Ala Thr Val Thr Ile Ala Lys Asn Ile Ser His His Leu Val
195             200             205

Val Val Ser Ile Asp Leu Lys Ser Ile Ser Ile Lys Leu Thr Val Val
210             215             220

Lys Thr Asn Glu Pro Val Asn Val Asp
225             230

```

INFORMATION FOR SEQ ID NO:77:

SEQUENCE CHARACTERISTICS

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:78:

```

Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Glu Gln Ala Ala
1           5           10           15
Ala Glu Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val
20           25           30
Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile
35           40           45
Thr Tyr Arg Ile Lys Leu Glu Val Ser Phe Lys Met Arg Pro Ala Gln
50           55           60
Pro Arg
65

```

2 INFORMATION FOR SEQ ID NO:78:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 69 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

x1 SEQUENCE DESCRIPTION: SEQ ID NO:79:

```

Val Pro Pro Ala Pro Pro Leu Pro Pro Leu Pro Pro Ser Pro Ile Ser
1           5           10           15
Lys Ala Ser Pro Pro Ser Pro Pro Leu Pro Pro Ala Pro Pro Val Ala
20           25           30
Pro Ile Pro Pro Met Pro Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro
35           40           45
Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro
50           55           60

```

or Pro Pro Pro Pro

3 INFORMATION FOR SEQ ID NO:79:

1. SEQUENCE CHARACTERISTICS:
- A LENGTH: 66 amino acids
 - B TYPE: amino acid
 - C STRANDEDNESS: single
 - D TOPOLOGY: linear

Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala
 20 25 30
 Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu
 35 40 45
 Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val
 50 55 60
 Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr
 65 70 75 80
 Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
 85 90 95
 Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
 100 105 110
 Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
 115 120 125
 Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
 130 135 140
 Gly Gly Val Ala Val Gly Gln Pro Val Val Ala Met Gly Asn Ser Gly
 145 150 155 160
 Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
 165 170 175
 Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Gln Gly Thr
 180 185 190
 Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
 195 200 205
 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
 210 215 220
 Ala Ala Thr Asn Asn Phe Ile Leu Pro Gln Ile Gly Ser Gly Thr Ala
 225 230 235 240
 Ile Pro Ile Gly Gln Ala Val Ala Ile Ala Gly Gln Ile Arg Ser Gly
 245 250 255
 Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu
 260 265 270
 Gly Val Val Asp Asn Asn Gly Asn Ile Ala Arg Val Gln Arg Val Val
 275 280 285 290

305 310 315 320
 Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
 325 330 335
 Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
 340 345 350
 Pro Pro Ala
 355

(2) INFORMATION FOR SEQ ID NO:81:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 305 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(A1) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr
 1 5 10 15
 Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala
 20 25 30
 Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys
 35 40 45
 Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala
 50 55 60
 Asp Pro Leu Ala Ala Lys Gly Val Lys Thr Tyr Asn Asp Ser Gln Ile
 65 70 75 80
 Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp
 85 90 95
 Asp Thr Ser Asn Leu Thr Ser Gln Ser Gln Leu Ser Thr Ser Val Val
 100 105 110
 Ser Asn Pro Ala Val Gln Thr Asp Leu Leu Ser Gly Val Thr Asn
 115 120 125
 Leu Gln Ala Gln Gly Thr Gln Val Ile Asp Gly Ile Ser Thr Thr Lys
 130 135 140
 Ile Thr Ile Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly
 145 150 155 160

180

185

190

Leu Thr Gln Ser Lys Trp Asn Glu Pro Val Asn Val Asp
 195 200 205

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
 1 5 10 15
 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
 20 25 30
 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val
 35 40 45
 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
 50 55 60
 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
 65 70 75 80
 Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
 85 90 95
 Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
 100 105 110
 Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
 115 120 125
 Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Gly Tyr
 130 135 140
 Asp Thr Ala Ala Asp Ala Thr Ser Ala Thr Ser Thr Glu Ile Asp
 145 150 155 160
 Arg Asn Gly Thr Val Ile Glu Thr Glu Lys Leu Arg His Pro Asn Arg
 165 170 175
 Phe Ala Gly Thr Pro Tyr Ala Thr Arg Ala Leu Thr Asn Ala Arg Gly
 180 185 190

```

210                215                220
Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
225                230                235                240
Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
245                250                255
Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
260                265                270
Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys
275                280                285

```

2 INFORMATION FOR SEQ ID NO:83:

1. SEQUENCE CHARACTERISTICS:
 - A) LENGTH: 173 amino acids
 - B) TYPE: amino acid
 - C) STRANDEDNESS: single
 - D) TOPOLOGY: linear

(XXI) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```

Thr Lys Phe His Ala Leu Met Gln Gln Gln Ile His Asn Glu Phe Thr
1                5                10                15
Ala Ala Gln Gln Tyr Val Ala Ile Ala Val Tyr Phe Asp Ser Glu Asp
20                25                30
Leu Pro Gln Leu Ala Lys His Phe Tyr Ser Gln Ala Val Glu Glu Arg
35                40                45
Asn His Ala Met Met Leu Val His His Leu Leu Asp Arg Asp Leu Arg
50                55                60
Val His Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro
65                70                75                80
Glu Glu Val Lys Glu Leu Val Val Asn Ile His Val Thr Val Thr Asp
85                90                95
His Val Gly Arg Leu Thr Ala Val Ala Arg Asp Glu Gly Asn Phe Leu
100                105                110
Glu His Gln Ser Met His Trp Phe Leu His Glu Gln Ile Glu Glu Val
115                120                125
Ala Leu Met Ala Thr Leu Val Asn Val Ala Asn Arg Ala Gly Ala Asn
130                135                140

```

165

170

2. INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:84:

```

Arg Ala Asp Glu Arg Lys Asn Thr Thr Met Lys Met Val Lys Ser Ile
 1             5             10             15

Ala Ala Gly Leu Thr Ala Ala Ala Ala Ile Gly Ala Ala Ala Ala Gly
 20             25             30

Val Thr Ser Ile Met Ala Gly Gly Pro Val Val Tyr Gln Met Gln Pro
 35             40             45

Val Val Phe Gly Ala Pro Leu Pro Leu Asp Pro Xaa Ser Ala Pro Xaa
 50             55             60

Val Pro Thr Ala Ala Gln Trp Thr Xaa Leu Leu Asn Xaa Leu Xaa Asp
 65             70             75             90

Pro Asn Val Ser Phe Xaa Asn Lys Gly Ser Leu Val Glu Gly Gly Ile
 85             90             95

Gly Gly Xaa Glu Gly Xaa Xaa Arg Arg Xaa Gln
100             105

```

3. INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```

Val Leu Ser Val Arg Glu Thr Asn Glu Phe Pro Xaa Ala Val Val Asn
 1             5             10             15

Pro Leu Gly Gln Pro Ile Asp Gly Arg Gly Asp Val Asn Ser Asn Thr
 20             25             30

Arg Arg Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Arg Gln Gly
 35             40             45

```

65 70 75 80
 Gly Lys Asn Arg Arg Leu Cys Arg Thr Pro Ser Ser Asn Gln Arg Glu
 85 90 95
 Glu Leu Gly Val Arg Trp Ile Pro Arg Ser Arg Cys Ala Cys Val Tyr
 100 105 110
 Val Gly His Arg Ala Arg Arg Gly Thr Tyr His Arg Arg
 115 120 125

(2) INFORMATION FOR SEQ ID NO:86.

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 117 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:86:

Cys Asp Ala Val Met Gly Phe Leu Gly Gly Ala Gly Pro Leu Ala Val
 1 5 10 15
 Val Asp Gln Gln Leu Val Thr Arg Val Pro Gln Gly Trp Ser Phe Ala
 20 25 30
 Gln Ala Ala Ala Val Pro Val Val Phe Leu Thr Ala Trp Tyr Gly Leu
 35 40 45
 Ala Asp Leu Ala Glu Ile Lys Ala Gly Glu Ser Val Leu Ile His Ala
 50 55 60
 Gly Thr Gly Gly Val Gly Met Ala Ala Val Gln Leu Ala Arg Gln Trp
 65 70 75
 Gly Val Glu Val Phe Val Thr Ala Ser Arg Gly Lys Trp Asp Thr Leu
 80 85 90
 Arg Ala Lys Lys Phe Ala Val Ile Pro Thr Ala Val Leu Pro His Lys
 95 100 105
 Lys Ser Ser Lys Gly
 110

(3) INFORMATION FOR SEQ ID NO:87

1. SEQUENCE CHARACTERISTICS:
- A LENGTH: 111 amino acids
 - B TYPE: amino acid
 - C STRANDEDNESS: single

```

Met Tyr Arg Phe Ala Cys Arg Thr Leu Met Leu Ala Ala Cys Ile Leu
1           5           10           15
Ala Thr Gly Val Ala Gly Leu Gly Val Gly Ala Gln Ser Ala Ala Gln
20           25           30
Thr Ala Pro Val Pro Asp Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp
35           40           45
Pro Ala Trp Gly Pro Asn Trp Asp Pro Tyr Thr Cys His Asp Asp Phe
50           55           60
His Arg Asp Ser Asp Gly Pro Asp His Ser Arg Asp Tyr Pro Gly Pro
65           70           75           80
Ile Leu Glu Gly Pro Val Leu Asp Asp Pro Gly Ala Ala Pro Pro Pro
85           90           95
Pro Ala Ala Gly Gly Gly Ala
100

```

12. INFORMATION FOR SEQ ID NO:88:

12.1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

12.2 SEQUENCE DESCRIPTION: SEQ ID NO:88:

```

Met His Cys Arg Val Trp Leu Glu Ile Gln Trp Arg Gly Met Leu Gly
1           5           10           15
Ala Asp Gln Ala Arg Ala Gly Gly Ser Ala Arg Ile Trp Arg Gln His
20           25           30
Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Ile Ala
35           40           45
Thr Lys Lys Thr Arg Lys Ile Ala Met Arg Val Trp Ser Ala Gly His
50           55           60
Glu Arg Leu Val Ala Glu Leu Thr Pro Asp His Ala Ala Ala Leu Gly
65           70           75           80
Asp Gln Leu Lys Gly Val Thr Ser
85

```

13. INFORMATION FOR SEQ ID NO:89:

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:89:

```

Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile
1           5           10           15
Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
20           25           30
Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
35           40           45
Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu
50           55           60
Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
65           70           75           80
Ala Asp Glu Gln Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
85           90           95

```

(X2) INFORMATION FOR SEQ ID NO:90:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 166 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1           2           3           4           5           6           7           8           9           10
Arg Ala Asn Gln Val Gln Ala Pro Met Ala Asp Pro Pro Thr Asp Val
15           20           25           30
Asp Ile Thr Arg Thr Gln Leu Thr Thr Lys Lys Asn Ala Ala Gln Gln
35           40           45
Ala Val Ser Ser Ala Gln Asn Met Ala Thr Tyr Leu Ala Ala Gln Ala
50           55           60
Lys Gln Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Lys
65           70           75           80
Thr Gly Gln Val Asp Ile Gln Ala Ala Thr Ala Leu Asp Asn Asn Gly
85           90           95

```

Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
 130 135 140

Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Trp Asn Thr Xaa Thr
 145 150 155 160

Leu Thr Leu Gln Gly Asp
 165

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Arg Ala Glu Arg Met
 1 5

(3) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 263 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Val Ala Thr Met Ser Val Thr Ala Gly Gln Ala Gln Leu Thr Ala Ala
 1 5 10 15

Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr
 20 25 30

Ala Pro Arg Ser Val Gln Ala Thr Asn Arg Ala Thr Leu Met Thr Leu
 35 40 45

Ala Ala Thr Asn Val Thr Gly Thr Asn Thr Ser Ala Ala Ala Thr Asn
 50 55 60

Gln Ala Gln Tyr Gln Gln Met Trp Ala Thr Asp Ala Ala Ala Met Pro
 65 70 75 80

Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Thr Trp Phe
 85 90 95

```

115              120              125
Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly
130              135              140
Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro
145              150              155              160
His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met
165              170              175
Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met
180              185              190
Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr Ala
195              200              205
Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly
210              215              220
Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
225              230              235              240
Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
245              250              255
Arg Arg Asn Gly Gly Pro Ala
260

```

1. INFORMATION FOR SEQ ID NO:93:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 303 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:93:

```

Met Thr Thr Ser Pro Lys Asn Glu Gly Tyr Glu His Ala His Pro Ala
1      5      10      15      20
Val Ser Thr His Gly Val Thr Glu Ser Phe Ala His Ala Asp Glu Gly
21     25     30     35     40     45
Ala Ser Lys Leu His Met Tyr Leu Asn Ile Ala Val Ala Val Leu Gly
46     51     56     61     66     71
Leu Ala Ala Tyr Phe Ala Ser Thr Glu Pro Met Pro Thr Leu Ser Thr
76     81     86     91     96     101

```

35	90	95
Pro Lys Ala Lys Ser His Val Thr Val Val Ala Val Leu Gly Val Leu 100 105 110		
Gly Val Phe Leu Met Val Ser Ala Thr Phe Asn Lys Pro Ser Ala Tyr 115 120 125		
Ser Thr Gly Trp Ala Leu Trp Val Val Leu Ala Phe Ile Val Phe Gln 130 135 140		
Ala Val Ala Ala Val Leu Ala Leu Leu Val Glu Thr Gly Ala Ile Thr 145 150 155 160		
Ala Pro Ala Pro Arg Pro Lys Phe Asp Pro Tyr Gly Gln Tyr Gly Arg 165 170 175		
Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly 180 185 190		
Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln 195 200 205		
Gln Ser Pro Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser 210 215 220		
Ser Ser Pro Ser Gln Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala 225 230 235 240		
Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser 245 250 255		
Thr Pro Ser Thr Ile Phe Pro Ser Phe Ser Pro Pro Pro Pro Val Ser 260 265 270		
Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Pro Val Asn Tyr Ser Asn 275 280 285		
Pro Ser Gly Gly Gln Gln Ser Ser Ser Pro Gly Gly Ala Pro Val 290 295 300		

INFORMATION FOR SEQ ID NO 34

1. SEQUENCE CHARACTERISTICS

- A LENGTH: 307 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO: 34

```

GTCTTCGGCG CGGCACTGCC GTTGGACCCG GCATCCGCCC CTGACSTCCC GACCGCCGCC      180
CAGTTGACCA GCCTGCTCAA CAGCCTCGCC GATCCCAACG TGTCGTTTGC GAACAAGGGC      240
AGTCTGGTCC AGGGCGGCAT CGGGGGCACC GAGGCGCGCA TCGCCGACCA CAAGCTGAAG      300
AAGGCCGCGG AGCACGGGGA TCTGCCGCTG TCGTTCAGCG TGACGAACAT CCAGCCGGCG      360
GCCGCCGGTT CGGCCACCGC CGACGTTTCC GTCTCGGGTC CGAAGCTCTC CTCGCCGGTC      420
ACGCAGAACC TCACGTTCTG GAATCAAGGC GGCTGGATGC TGTACCCGC ATCGGCGATG      480
GAGTTGCTCC AGGCCGCGG CAACTGA                                          507

```

2) INFORMATION FOR SEQ ID NO:95:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 168 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

2) SEQUENCE DESCRIPTION: SEQ ID NO:95:

```

Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala Ala Ala Ala
1      5      10      15
Ile Gly Ala Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro
20     25     30
Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro Leu Pro Leu
35     40     45
Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
50     55     60
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn Lys Gly
65     70     75     80
Gln Leu Val Val Val Val Val Val Val Thr Val Ala Ala Ile Ala Asp
85     90     95     100
His Lys Leu Lys Gln Ala Ala Val His His Asp Leu Pro Leu Ser Asp
105    110    115
Ser Val Thr Asn Ile Ser Ser Ala Ala Ala Thr Ser Ala Thr Ala Asp
120    125    130
Ala Ser Val Ser Val Pro Val Leu Ser Ser Val Val Thr Val Asn Val
135    140    145

```

Glu Leu Leu Gln Ala Ala Gly Asn
165

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CGTGGCAATG TCSTTGACCG TCGGGGCCGG GGTGGGCTCC GCAGATCCCG TGGACGCGGT	60
CATTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCC CTCAACCCGA CGGATCCGGG	120
GGCTCCCGCA CAGTTCAACG COTCACCAGT GGGCAGTCC TATTTGCCCA ATTTCTCTCC	180
TGCACGGCCA COTCAGCCCG CTGCCATGGC GGGCAATTG CAAGCTGTCC CCGGGGCGGC	240
AGAGTAGATC GGCCTTCTCC AGTCGGTTGC CGGCTCTTCC AAGAACTATT AAGCCCATGC	300
GGGCCCCATC GCGCGACCCG GCATCGTCCG CCGGGCTAGG CCAGATTGCC CCGCTCTCA	360
ACGGGCGCCA TCCCGCAGC CGGCATCTGC GCGGGGGCTA GGCCAGATTG CCGGCTCTCT	420
CAACGGGCGG CATCTCTTCC CGAATTCTTG CAGCCCGGGG GATCCACTAG TTCTAGAGCG	480
GGCGCGACCG GGGTGGAGGT	500

(3) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
 A. LENGTH: 96 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Val Ala Met Ser Leu Thr Val Val Ala Gly Val Ala Ser Ala Ser Pro	10	20	30
Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val Val Ala	40	50	60
Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser	70	80	90
Pro Val Ala Gln Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr	96		

Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr
 85 90 95

(2) INFORMATION FOR SEQ ID NO:98:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 154 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:98:

ATGACAGAGC AGCAGTGGAA TTTCGGGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA 60
 AATGTACAGT CCATTGATTC GGTCTTGAC GAGGGGAAGC AGTCCCTGAC CAGCTGGCA 120
 GCGGCTGGG GCGGTAGCCG TTGGGAAGCG TACC 154

(2) INFORMATION FOR SEQ ID NO:99:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Met Thr Glu Phe Phe Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
 1 5 10 15
 Ala Ile Glu Phe Asn Val Thr Ser Ile His Ser Leu Leu Asn Gly Gly
 21 25 30
 Lys Glu Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
 35 40 45
 Ala Ala Tyr
 51

(2) INFORMATION FOR SEQ ID NO:100:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GCTCGAAACG CGGCACAGCC GACGGTGGCT CCGNCGAGGC GCTGNCCTCA AAATCGTTGA 180
 GACAATTCCN CGGGGGGCGC TACAAGGAAG TCGGTGCTGA ATTGNCBNG TATCTGCTCG 240
 ACCTGTGTGG TCTGNAGCCG GACGAAGCGG TGCTCGACGT CG 282

(2) INFORMATION FOR SEQ ID NO:101:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:101.

GATGCTACCC GTGCGAGTGC TCGGGCCCTT TGAGGATGGA GTGCACCTGT CTTCCTGAT 50
 GGCATACCCA GAGATGTTGG CGGCGGCGCC TGACACCTG CAGAGCATCG GTGCTACCA 120
 TGTGCTAGC AATGCCCTG CGGCGGCGCC GAGCACTGGG GTGCTGCGCC CGGCTGCGCA 180
 TGAGGTGTGG GCGCTGACTG CGGCGCACTT CGCGGCACAT CGCGCGATGT ATCAGTCTGT 240
 GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGCGC ACCCTTCCCA GCACCGCCAG 300
 GTGCTATGCG GCGCTGAAG TCGCCAATGC GCGCGCGCGC AGCTAAGCCA GGAACACTCG 360
 GCACGAGAAA CCACGAGAAA TAGGGACACG TAATGCTGGA TTTCGGGCGC TTACCACTGG 420
 AGATGAATC CGCGAGGATG TACCGCGCGC CGGCTTGGCG GTGCTGCTG CGCGCGCTG 480
 AATGCTGGA GAGCTGCGCG ATGACCTGT TTTCGCGCGC TTGCGCTTT TGTGCTGCG 540
 GTGCGCTGT GACGCTGCGC TCGTGATAG GTTCGTGCGC CGCTTGTATG GTGCGCGCGC 600
 GCTGCGCTGA TGTGCGCTGG ATGAGCTGCA TCGCGCGCGC CGCGGAGCTG ACCGCGCTG 660
 GTTCGCTGT TGTGCGCGCG TGTACGAGA TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 720
 TGTGCTGCTG GAGCTGCTGT GAGCTGCTG TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 780
 TGTGCTGCTG GAGCTGCTGT GAGCTGCTG TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 840
 TGTGCTGCTG GAGCTGCTGT GAGCTGCTG TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 900
 TGTGCTGCTG GAGCTGCTGT GAGCTGCTG TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 960
 TGTGCTGCTG GAGCTGCTGT GAGCTGCTG TGTGCTGCTG TGTGCTGCTG TGTGCTGCTG 1020

CGATGACCAA CTCGGGTGTG TCGATGACCA ACACCTTGAG CTCGATGTTG AAGGGCTTTG	1200
CTCCGGCGGC GCGCGCCAG GCCGTGCAAA CCGCGGCGCA AAACGGGGTC CGGGCGATGA	1260
GCTCGGTGGG CAGCTGCTG GTTCTTTGGG GTCTGGGCGG TGGGGTGGCC GCCAACTTGG	1320
GTGGGGCGGC CTCGGTGGGT TCGTTGTGGG TCCCGCAGGC CTGGGCGCGG GCCAACCAGG	1380
CAGTCACCCC GCGCGCGCGG GCGCTGCGGC TGACCAGGCT GACCAGCGCC GCGGAAAGAG	1440
GGCCCGGGCA GATGCTGGGC GGGCTGCCGG TGGGGCAGAT GGGCGCCAGG GCCGGTGGTG	1500
GGGTCACTGG TGTGCTGCGT GTTCCGCGCG GACCTTATGT GATGCCGCAT TCTCCGGCGG	1560
CGGGCTAGGA GAGGGGCGCG AGACTGTGCT TATTTGACCA GTGATCGGCG GTCTCGGTGT	1620
TTCCCGCGCG CGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG	1680
TTTCAACAAG GAGACAGGCA ACATGGCGTC AGTTTTATG ACCGATCCCG ACCCGATCGG	1740
GGACATGGCG GCGCGTTTTG AGGTGCACCG CGAGACGGTG GAGGACGAGG CTCGCCGAT	1800
GTGGCGCTCC GCGCAAAACA TTTCCGGTCC GGGGTGGAGT GGGATGGCGG AGGCGACCTC	1860
GCTAGACACC ATGGCCGAGA TGAATCAGCG GTTTCGCAAC ATGCTGAACA TGCTGCACGG	1920
GCTCGCTGAC CGGCTGGTTG CGGACGCCAA CAACTACGAG TAGCAAGAGC AGGCCTCCCA	1980
GCACATGCTC AGCAGCTAAC CTCAGCGGCT GCAGGACAAT ACTTTTACAA GCGAAGGAGA	2040
AGAGGTTTGA TGACGATCAA GTATCAATTC GGGGATGTCC ACCGTACCGG CGCCATGATC	2100
TGCGTGAAGG TGCGGTGCTT CGAGCGCGAG CATCAGGCCA TCATTGCTGA TGTGTTGAGC	2160
TGCACTGACT TTGCGCGCGG CGCGCGTTCC CGGGCTGCG AGGGGTTGAT TACCCAGTTG	2220
TGCTGTAAT TGCACTGAT CTACGACGAG CGCAACGCGC ACGGGCAGAA GCTGCAGGCT	2280
TGCGCAACA AGATGAGCA ACCGACAGCT TCGTGGGCT TCGCTGGGC CTGACACCAG	2340
TGAACTGA TGAATGCT TGAAGTGA ACTTCTGCT TGATGTTT TGTGCGCAAT	2400
TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT	2460
TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT	2520
TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT	2580
TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT	2640
TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT TGAATGCT	2700

```

GGTGGCCACC CACGGCCAGG AGGGCTTCGG GGTGGCTGCC ATCAGATTGG CTGGCTAGTG      2880
GTTTCTGCAG CGCTGCCAGG CCGCTGCCGG CAGGGTGGCG CCGATCGCGG CCACCAGGCC      2940
GGCGTGGGCG TCGTGGTGA CCAGCCCGAC CCGGACAGG CCGCGGGCGA CCAGGTCCGC      3000
GAAGAACGCC AGCCAGCCGG CCCCCTCCTC GCGGAGGTG ACCTGGATGC CCAGGATC      3058

```

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met      1
1           5           10           15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp      20
20          25          30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser      35
35          40          45
Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly      50
50          55          60
Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr      65
65          70          75          80
Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala      85
85          90          95
Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala      100
100         105         110
Glu Asn Arg Ala Ile Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Ile      115
115         120         125
Gln Asn Thr Pro Ala Ile Ala Val Asn Ile Ala Gln Thr Ile Thr Met      130
130         135         140
Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala      145
145         150         155         160
Thr Ala Thr Ala Thr Leu Leu Phe Phe Ile Gln Ala Pro His Met Thr      165
165         170         175

```

Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn
 225 230 235 240
 Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val
 245 250 255
 Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala
 260 265 270
 Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala
 275 280 285
 Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly
 290 295 300
 Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val
 305 310 315 320
 Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg
 325 330 335
 Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly
 340 345 350
 Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly
 355 360 365
 Gly Leu Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met
 370 375 380
 Pro His Ser Pro Ala Ala Gly
 385 390

INFORMATION FOR SEQ ID NO.111

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 1025 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO.111

CACTGATGCA GCGGCGCTGC AGGCTTCGAG CGTGGCTGCT TTTGATCTTC GGTGAGTTC

CCGTGTGGGG TCGATTGGG CGGACCAGTC CTCACCAAGC CTTGGCGTGC GCGCCAGGCG 240
 GCGGATCAGA TCGCTTGACT ACCAATCAAT CTTGAGCTCC CGGGCCGATG CTCGGGCTAA 300
 ATGAGGAGGA GCACCGCTGT CTTTCACTGC GCAACCGGAG ATGTTGGCGG CCGCGGCTGG 360
 CGAACTTCGT TCCCTGGGGG CAACGCTGAA CGCTAGCAAT GCGCGCCGAG CCGTGCCGAC 420
 GACTGCGGTG CTGCCCCCGG CTGCGGACGA GGTGTCGCTG CTGCTTGCCA CACAATTCCG 480
 TACGCATGCG GCGACGTATC AGACGGCCAG CGCCAAGGCC GCGGTGATCC ATGAGCAGTT 540
 TGTGACCAGC CTGCGCAACA GCGCTAGTTC ATATGCGGAC ACCGAGGCCG CCAACCGTGT 600
 GGTGACCGGC TAGCTGACCT GACCGTATTC GAGCGGAAGG ATTATCGAAG TGGTGGATTT 660
 CCGGGCGTTA CCACTCGAGA TCACTCCCG GAGGATGTAC GCGGCGGCGG ATTGCGGCTC 720
 GCTGCTGGCC GCGCGGAAGA TGTGGGACAG TGTGGCGAGT GACCTGTTTT GCGCGCGCTC 780
 GCGCTTTTCAG TCGGTGCTGT GGGGTCTGAC CGTGCGGCTC TCGATAGGTT TGTGCGCGCG 840
 TGTGATGGCG GCGCGCGGCT CGCGTATGT GCGCTGGATG AGCGTCACCG GCGGCGAGGC 900
 CGAGCTGACC GCGCGCCAGG TCGGGGTTGC TCGCGCGGCC TACGAGACAG CGTATAGGCT 960
 GACGGTGGCC CCGCGCGTGA TCGCGAGAA CGCTACCGAA CTGATGACCG TGACCGCGAC 1020
 GAACCTCTTG GGGCAAAACA CCGCGCGGAT CGAGCGCAAT CAGGCGCGCAT ACAGCCAGAT 1080
 GTGCGCGCA GACCGCGAGG CCACTATCG CTACGCGCGC ACGGCGCGCA CCGCGACGAA 1140
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1200
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1260
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1320
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1380
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1440
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1500
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1560
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1620
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1680
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1740
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1800
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1860
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1920
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 1980
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2040
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2100
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2160
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2220
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2280
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2340
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2400
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2460
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2520
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2580
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2640
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2700
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2760
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2820
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2880
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 2940
 GCGCTTCTG CCGTTGAGG AGCGCGGCT GATGACCAAC CCGCGCGCGC TCGTTGAGCA 3000

(A) LENGTH: 159 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
1             5             10             15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp
20             25             30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
35             40             45
Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
50             55             60
Leu Met Ala Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
65             70             75             80
Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
85             90             95
Ala Tyr Glu Thr Ala Tyr Arg Leu Thr Val Pro Pro Pro Val Ile Ala
100            105            110
Gln Asn Arg Thr Gln Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly
115            120            125
Gln Asn Thr Pro Ala Ile Glu Ala Asn Gln Ala Ala Tyr Ser Gln Met
130            135            140
Pro Gly Gln Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala
145            150            155            160
Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr
165            170            175
Asn Pro Gly Gly Ser Ser Val Ala Ala Val Ala Val Gln Ile Ala Ile
180            185            190
Asp Thr Ala Ala Ala Asn Ile Leu Met Asn Asn Val Pro Gln Ala Leu
195            200            205
Gln Gln Leu Ala Gln Pro Ala Gln Gly Val Val Pro Ser Ser Lys Leu
210            215            220
Gln Gly Leu Thr Thr Ala Val Ser Ser Thr Leu Ser Pro Leu Ser Asn
225            230

```

Ser Met Thr Asn Thr Leu His Ser Met Leu Lys Gly Leu Ala Pro Ala
 260 265 270

Ala Ala Gln Ala Val Glu Thr Ala Ala Glu Asn Gly Val Trp Ala Met
 275 280 285

Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu
 290 295 300

Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser
 305 310 315 320

Leu Ser Val Pro Pro Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro
 325 330 335

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr
 340 345 350

Ala Pro Gly His Met Leu Gly
 355

1. INFORMATION FOR SEQ ID NO:105:

1.1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1027 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

1.2 SEQUENCE DESCRIPTION: SEQ ID NO:105:

ATTTCAGTGG AGAATGATAT TGAAGGCGCTG TATCCACGGAT GGCTGAGACA ACCGAACCAAT 60

TCTGGGACCG GGGGACATCG GAAGGCGAGG CGATGCGGTT GCGCGCGGAA GCGGAAGCGG 120

GCGAAGCGGA AGCGCTGGCG GCGCGCGCGG GCGCGCGTGC CGCTGCGCGG CGCTTGAAGC 180

GTGAGGCGCT GCGGATGGCT GAGGCGGAGG ACCAGAACCT GCGCGAGGAT ATGCAGACTG 240

GAAAGAGTTC GAAGACTATG AGAATATATG GAATATGAG TGTGAGAGCG AGGAGGCGCG 300

ATTTCAGTGG TCTGGGACCG GTTGTGTCGA ATTTCAGTTC GAAGAGATTC GAGGATTCAT 360

ATTTCAGTGG GAAGTGTGTA TATGTCGTA GTTTCAGTGG GTGAGGAGAT ACATTGTGTA 420

GAAAGAGTTC GAGGCGGAGG AAGGCGGAGG GTGCGCGCGG GCGTTCGCGG GCGGAGCGAA 480

GGAAGGTGTC ATGAACATGA GCTGCGTGGG GTTCAACAAG GCGAAGAGAG AGCTGCGCGG 540

ATTTCAGTGG AGTTCAGTGG GCGAATTCAG GTATGATTCG GAGGAGTGGG GAGCGGATTC 600

TTCCGCTGGG GCGAAAGACG AACCACTGC GTGGCGGGCTC AAAGTGACCG TGACCGAAGA	780
GGGGGGACAG TACAAGATGT CGAAAGTTGA GTTCGTACCG TGACCGATGA CTACCGGAC	840
GTCAACACCG AAACCACTGA CGCCACCGAA GTCGCTGAGA TCGACTCAGC CGCAGGCGAA	900
GCGGGTGATT CGGCGACCGA GGCATTTGAC ACCGACTCTG CAACGGAATC TACCGCCAC	960
AAGGGTCAGC GGCACCGTGA CCTGTGGCGA ATGCAGGTTA CCTTGAAACC CGTTCCGGTG	1020
ATTCTCATCC TGCTCATGTT GATCTCTGGG GGCGCGACCG GATGGCTATA CCTTGAGCAA	1080
TACGACCGCA TCAGCAGACG GACTCCGGCG CGCGCCGTGC TGCCGTCGCG GCGGCGTCTC	1140
ACGGGACAAT CGCGCTGTTG TGTATTACCG CGACACCTCG ACCAAGACTT CGCTACCGCC	1200
AGGTCCGACC TCGCGCGCGA TTTGCTGTCC TATACGACCA GTTCAGCGAG CAGATCTGG	1260
GTCCGGCGGG CAAACAGAAG TCACTGAAAA CGACCGCGAA GTGGGTGCGC CGCGCCGTGT	1320
CGGAGCTACA TCGGATTCCG GCGCTGCTTC TGGTTTTTGT CGACGAGAGC ACTACCACTA	1380
AGGACAGCGC CAATCCGTCG ATGGCGGGCA CGAGCGTGAT GGTGACCTTA GCGAAGCTCG	1440
ACGGCAATTC GGTGATCACC AAGTTCACCC CGTTTATAGT TGCGGTAGGC GGTGCGCAAG	1500
TCTGACGGGG CGCGGGGTGG CTGCTGCTGC GAGATACCGG CGTTCTCTCG GACAATCAGC	1560
CGCCGACCTT AAACAGATCT CGGCGGCTGT CTAATCGGCG GGTATTATTA AGATTAGTTG	1620
CGACTGATTT TACCTGATCT TCAGATTGTT CAGCTGGATT TAGCTTCGCG CGAGGGCGGG	1680
TGCTGCACTT TGCATCTCGG TTTGTGACTA GTTGAGAGAA TTTGACCTGT TCGGACCTT	1740
TTTTGTCTG CATGATTCGT CCTAGTTATG CGCGAGCGCA AGGATTATCG AAGTGTGGA	1800
TTTGGGGGG TTACGACCGG AGATCAACTC CGCGAGGATG TACCGCTGCG CGGTTCTGCG	1860
TTTGTGCTG TCGCGCGCGA AGATCTGCGA CAGCTGCGCG AGTGAATCTT TTTCGCGCTT	1920
TTTATTTT CATGCTGCG TTTGGGTGT TACGAGCGA TCGCGATAG TTGCTGCGCG	1980
TTTATTTA TCGCGCGCG AGTCTGCTA GTTGTGCTA ATGATTTGA TTGCGCGCGA	2040
CGTGAATCTT ATCGCGCGCG AGTCTGCTT TCGTCTGCGG CGCTAGAGA CGCGCTATCG	2100
CTGAATCTT TCGCGCGCG TATCGCGCGA GAGCTGCTT GAAATGATTA TTCTGATAGC	2160
TATGAATCTT TCGCGCGCG ATA TCGCGCGA TATGCGCTT GAAATGATTA TTCTGATAGC	2220
TATGAATCTT TCGCGCGCG ATA TCGCGCGA TATGCGCTT GAAATGATTA TTCTGATAGC	2280

```

TGTGCCCCAA GCGCTGCAAC AACTGGCCCA GCGCACGAAA AGCATCTGGC CGTTCGACCA   2460
ACTGAGTGAA CTCTGGAAAG CCATCTCGCC CCATCTGTCC CCGCTCAGCA ACATCGTGTC   2520
GATGCTCAAC AACCACGTGT CGATGACCAA CTCGGGTGTG TCGATGGCCA GCACCTTGCA   2580
CTCAATGTTG AAGGGCTTTG CTCGGGCGGC GGCTCAGGCC GTGGAAACCG CGGCGCAAAA   2640
CGGGGTCCAG GCGATGAGCT CGCTGGGCAG CCAGCTGGGT TCGTCGCTGG GTTCTTCGGG   2700
TCTGGGCGCT GGGGTGGCCG CCAACTTGGG TCGGGCGGCC TCGGTCGGTT CGTTGTGGGT   2760
GCCCCAGGCC TGGGCCCGCG CCAACCAGGC GGTACCCCCG GCGGCGCGGG CGCTGCCGCT   2820
GACCAGCGTG ACCAGCGCCG CCAAAACCGC CCCCCGACAC ATGCTGGGCG GGCTACCGCT   2880
CGGGCAACTG ACCAATAGCG CCGCGCGGTT CGGCGGGGTT AGCAATCGCT TCGGATGCC   2940
CGCGCGGGCG TACGTAATCG CCGGTGTGCC CCGCGCGGGG TAACGCGGAT CCGCACGCAA   3000
TCGCGGCGCT CTATCGCGCG ACGGATC                                     3027

```

2. INFORMATION FOR SEQ ID NO:106:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 396 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:106:

```

Val Val Asp Phe Ile Ala Met Thr Arg Ile Ile Asn Ser Ala Arg Met
1      15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp
21     25     30
Ser Ser Met Ala Ser Asp Ser Thr Ser Ala Ala Ser Ala Phe Ile Ser
36     40
Val Val Thr Phe Thr Thr Thr Ile Thr Thr Ile Gly Ser Ser Ala Ile
46     51
Leu Met Val Ala Ala Ala Ser Arg Thr Val Ala Trp Met Ser Val Thr
56     60     65     70     75     80
Ala Ile Ile Ala Ile Ser Thr Ala Ala Ile Val Arg Val Ala Ala Ala
86     91     96

```

115	120	125
Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met		
130	135	140
Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Thr Ala Ala		
145	150	155
Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr		
165	170	175
Asn Pro Gly Gly Leu Leu Glu Gln Ala Val Ala Val Glu Glu Ala Ile		
180	185	190
Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu		
195	200	205
Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu		
210	215	220
Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn		
225	230	235
Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val		
245	250	255
Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala		
260	265	270
Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met		
275	280	285
Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu		
290	295	300
Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser		
305	310	315
Leu Ser Val Pro Thr Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro		
320	325	330
Ala Ala Asn Ala Leu Thr Leu Thr Ser Leu Thr Ser Ala Val Gln Thr		
335	340	345
Ala Pro Gly His Met Leu Gly Gly Leu Pro Leu Gly Gln Leu Thr Asn		
350	355	360
Ser Gly Gly Gly Phe Gly Gly Val Ser Asn Ala Leu Arg Met Pro Pro		
365	370	375
Asn Ala Thr Val Met Ser Ala Val Pro Ala Ala Ile		

(i) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 1616 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

```

CATCGGAGGG AGTGATCACC ATGCTGTGGC ACGCAATGCC ACCGGAGTAA ATACCGCACG      60
GCTGATGGCC GCGCGGGGTC CGGCTCCAAT GCTTGCGGCG GCCGCGGGAT GGCAGACGCT      120
TTGCGCGGCT CTGGACGCTC AGGCCGTCCA GTTGACCGCG CGCCTGAACT CTCTGGGAGA      180
AGGCTGGACT GGAGGTGGCA GCGACAAGCC GCTTGCGGCT GCAACGCCCA TGGTGGTCTG      240
GCTACAAACC GCGTCAACAC AGGCCAAGAC GCTTGGATG CAGGCGAGCC GCGAAGCCCG      300
GGCATAACCT CAGGCGATGG GACGAGCCCG GTGCGTCCCG GAGATCGCCG CCAACCCACAT      360
CACTCCAGGG GTGCTTACCG CCACCAACTT GTTGGGTATC AACAGCATCC CGATCGCGTT      420
GACCGAGATG GATTATTTCA TGGTATGTG GAACAGGCA GCGCTGGCAA TGGAGGTCTA      480
CGAGGCCGAG ACCGCGGTTA ACACGCTTTT CGAGAAGCTC GAGCGCATGG COTCGATCCT      540
TGATCCCGGC GCGAGCCAGA GCACGACGAA CCGGATCTTC GGAATGCCCT CCGCTGGCAG      600
GTCAACACCG GTTGGCGAST TCGCGCGGCG GGTACCGAG ACCTGCGGCC AACTGGGTGA      660
GATGAGCGGC CGCATCGAGC AGCTGACGCA TCGGCTCCAG CAGGTGACCT GCTTGTTCAG      720
TGAGGTGGCG GCGACCGCCG CGGCGAAGCC AGTTGACGAG GAAGCGCGCG AGATGGGCTT      780
GCTCGGCAAC AGTCCGCTGT CGAAGCATCC GTTGGTCTGT CGATCAGGCC CGAGCGCGCG      840
CGCGCGCGCT GTGCGCGCGG AGTCCGTACC TCGGCGAGGT GGTGCTTGA CGCGACCGCG      900
GCTGATCTCT CAGCTGATCG AAAAGCGCGT TCGGCTCG GTGATCGCG CGCTGCTGTG      960
GAAAGCTG CCAAGGCT TCGCGCTG GTTCTGCG GAATGATG GTGATCTG      1020
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1080
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1140
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1200
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1260
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1320
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1380
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1440
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1500
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1560
GAAAGCTG CAGTCAAGCA CAGGCTGT GTTCTGCG GAATGATG GTGATCTG      1616

```

AATATTCGTC AGGCCCGGCT GCAATACTCG AGGCCCGGACG AGGAGCAGCA GCAGGCCGCTG 1500
 TCCTCGCAAA TGGGCTTCTG ACCCGCTAAT ACGAAAAGAA ACGGAGCAAA AACATGACAG 1560
 AGCAGCAGTG GAATTTTCGGG GGTATCGAGG CCGCGGCAAG CGCAATCCAG GGAAAT 1616

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 432 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CTASTGGATG GGACCATGGC CATTTTCTGC AGTCCTACTG CTTCTGTGTG TGACATTTTG 60
 GCACGCCGCG GCAGAACGAAG CACTGGGGTC GAAGAACGGC TGCCTGCCA TATCTCCGG 120
 AGTTTCATA CTTCTGTGCG GCGGGAAGAG CTTCTGTAG TCGGCCGCA TGACAACCTC 180
 TCAGAGTGGC CTCAAACGTA TAAACACGAG AAAGGCGAG ACCACGGAA GGTGGAATC 240
 GCGGATCCG CTGTTTGGCT ATTCTACGCG AACTCGGCGT TGGCTATGC GAACATCCCA 300
 CTCAGCTTGC CTTCTGTGCA AGCCATTGCC TGACCGGCTT CGCTGATCCT CCGGCCGAGG 360
 TTCTGAGCG CTTCTGTGAG CTGGGTAGCG CTGGCTGCC ATTTTTCCTG GACAGCTCTG 420
 TTAGCTCTCG AA 480

(iii) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 368 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Met Leu Trp His Ala Met Ser Ser Cys Ala Asn Thr Ala Arg Ser Met
 1 10 19
 Val Gly Ala Gly Pro Ala Pro Met Leu Ala Ala Ala Ala Gly Trp Gly
 21 30 39 48
 Ser Leu Ser Ala Ala Leu Thr Ala Thr Ala Gly Thr Thr Thr Arg
 59 68 77 86 95 104 113 122 131 140 149 158 167 176 185 194 203 212 221 230 239 248 257 266 275 284 293 302 311 320 329 338 347 356 365 374 383 392 401 410 419 428 437 446 455 464 473 482 491 500 509 518 527 536 545 554 563 572 581 590 599 608 617 626 635 644 653 662 671 680 689 698 707 716 725 734 743 752 761 770 779 788 797 806 815 824 833 842 851 860 869 878 887 896 905 914 923 932 941 950 959 968 977 986 995 1004 1013 1022 1031 1040 1049 1058 1067 1076 1085 1094 1103 1112 1121 1130 1139 1148 1157 1166 1175 1184 1193 1202 1211 1220 1229 1238 1247 1256 1265 1274 1283 1292 1301 1310 1319 1328 1337 1346 1355 1364 1373 1382 1391 1400 1409 1418 1427 1436 1445 1454 1463 1472 1481 1490 1500

Leu Ala Ala Ala Thr Pro Met Val Val Trp Leu Gln Thr Ala Ser Thr
 65 70 75 80
 Gln Ala Lys Thr Arg Ala Met Gln Ala Thr Ala Gln Ala Ala Ala Tyr
 85 90 95
 Thr Gln Ala Met Ala Thr Thr Pro Ser Leu Pro Glu Ile Ala Ala Asn
 100 105 110
 His Ile Thr Gln Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn
 115 120 125
 Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp
 130 135 140
 Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Glu Thr Ala Val
 145 150 155 160
 Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro
 165 170 175
 Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro
 180 185 190
 Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr
 195 200 205
 Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln
 210 215 220
 Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly
 225 230 235 240
 Thr Gly Asn Pro Ala Asp Glu Gln Ala Ala Gln Met Gln Leu Leu Gly
 245 250 255
 Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser
 260 265 270
 Thr Ser Ala Gln Leu Leu Asn Ala Thr Ser Leu Pro Gly Ala Gln Gly
 275 280 285
 Thr Leu Thr Ala Thr Pro Leu Met Ser Gln Leu Ile Gln Lys Pro Val
 290 295 300
 Ala Pro Ser Val Met Pro Ala Ala Ala Ala Gly Ser Ser Ala Thr Gly
 305 310 315 320
 Thr Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gln Ala Thr Ser
 325 330 335

355

360

365

2. INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```

Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly
1           5           10           15

Asn Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val
20           25           30

Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly
35           40           45

Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys
50           55           60

Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly
65           70           75           80

Val Gln Tyr Ser Arg Ala Asp Glu Gln Gln Gln Gln Ala Leu Ser Ser
85           90           95

Gln Met Gly Phe
100

```

3. INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 396 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:111:

```

GATGTCGGGG GAGCTGAAAA GCGAGATCGA CAGCTGGAG TCGAGGCGAG GTTGGTTGCA      60
AGGCACTGG CCGCGCGCGG CCGCGAGCGG CCGCGAGCGG CCGCTGCTGG GTTGGGAGA      120
AGGAGCAAT AAGGAGAAG AGGAGCTGGA CAGATCTGG AGCAATATT GTGAGGGGAG      180

```

(2) INFORMATION FOR SEQ ID NO:112:

(A) LENGTH: 80 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala
 1 5 10 15
 Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
 20 25 30
 Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu
 35 40 45
 Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser
 50 55 60
 Arg Ala Asp Glu Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
 65 70 75 80

1 LENGTH: 387 base pairs
2 TYPE: nucleic acid
3 STRANDEDNESS: single
4 TOPOLOGY: linear

[illegible]

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:114:

```

CGGCACGAGG ATCTCGGTTG CCCCAACGGC GCTGGCGAGG GCTCCGTTCC GGGGGCGAGC      60
TGCCGCCCGG ATGCTTCCTC TGCCCGCAGC CGCGCCTGGA TGGATGGACC AGTTGCTACC      120
TTCCCGACST TTCGTTGGGT GTCTGTCCGA TAGCGGTGAC CCCGGCGGCG ACGTCGGGAG      180
TGTGGGGGGG CAGGCCGGGT CGGTGGTTCC GCGGGGAGCG CAGACGGTCT GGACGGAACS      240
GGCGGGGGTT CCGCGATTGG CATCTTTGCC CA                                     300

```

2. INFORMATION FOR SEQ ID NO:115:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:115:

```

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val
 1             5             10             15
Val Ala Ala Leu
          20

```

3. INFORMATION FOR SEQ ID NO:116:

(1) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 15 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:116:

```

Ala Val Gln Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
 1             5             10             15

```

4. INFORMATION FOR SEQ ID NO:117:

(1) SEQUENCE CHARACTERISTICS:

- A. LENGTH:
- B. TYPE:
- C. STRANDEDNESS:
- D. TOPOLOGY:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys
1 5 10 15

Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:118:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:119:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Asp Ile Gly Ser Glu Ser Thr Ile Asp Gln Gln Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:120:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro
1 5 10 15

Ser

(2) INFORMATION FOR SEQ ID NO:122:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:123:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
1 5 10 15

Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn
20 25 30

(2) INFORMATION FOR SEQ ID NO:124:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp Pro Pro Asp Pro His Glu Xaa Asp Met Thr Lys Gly Tyr Thr Pro
1 5 10 15

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Asp Pro Gly Tyr Thr Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:126:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Xaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr
1 5 10

(3) INFORMATION FOR SEQ ID NO:127:

(1) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 5 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: note= "The Third Residue Can Be Either a Gln or Leu"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Asn Pro Xaa Val Thr Ala Tyr Ala G
1 5

(4) INFORMATION FOR SEQ ID NO:128:

(1) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 5 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:

Xaa Xaa Xaa Glu Lys Pro Phe Leu Arg
1 5

(2) INFORMATION FOR SEQ ID NO:129:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Xaa Asp Ser Glu Lys Ser Ala Thr Ile Lys Val Thr Asp Ala Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:130:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Ala Gly Asp Thr Xaa Ile Tyr Ile Val Gly Asn Leu Thr Ala Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:131:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Ala Ser Ile Ser Gly Ala Gly Ser Gly Ala Thr Ala Ala Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:132:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Asn Val His Leu Val

20

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 882 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

```

GCAACGCTGT CGTGGCCTTT CCGGTGATCG GTTCCGCTC GCTGGCGGTG CCGGTGCGCG      60
TCACCATCGG ACCGACCGCG CCGTCAAAAC CCGTAGAGCG ACACAAAAAC CCGCAGCGAG      120
GGAACTTCAT CCGCTTCTTG CCGACCGAAC AGCAGGCGCG CCGTCCCGCG CCGTCCCGCG      180
ATGATCCGAC CCGTGGATTC CAGGCGCGGA CCGTCCCGCG TGTACGAAAC GTGCTGCTCG      240
CGCCCGGCTAC CTCACCGCGG GTGGGTGGGA CCGCGGCTTC GCGTGCCTCG GAAGCGCGCG      300
CCGTCCCGCG TGTCTGCTGT GCGCGGCTCG CAATCCCGGT CCGGATCCTC ATTCCCGCGT      360
TCCCGGCTTG GCAGCGTGGG ATCGCGAGCA TCGCGACCGG ATCGCGACCG ACCCGGCTGA      420
CGACGTGGGG GACGACCGCG CCGACCGCGG CGCGACCGAC TCGGCTGAGG ACCCGGCGAA      480
CGACCGCGCG GACGACCGTG CCGACCGCGG TCGCAAGGAC CCGCGCGACG ACCCGGCTGA      540
CGCGCGCGAG AACGACGCTG CCGCGAGGCA CCGTCCCGCG GACGACCGTG CCGCGCGAGG      600
CGTCCCGCGG GACGACCGTG CCGCGAGGCA CCGCGACCGG GACGACCGTG CCGCGCGAGG      660
CGCGCGCGCA CCGCGCGCA CCGCGACCGG AACGATGCG AACCGACGAG CAGACCGCTG      720
CGCGCGCGAG CCGCGCGCG CCGCGCGAG CCGCGCGCG CCGCGCGCG CCGCGCGCG      780
CGCGCGCGAG CCGCGCGCG CCGCGCGAG CCGCGCGCG CCGCGCGCG CCGCGCGCG      840
CGCGCGCGAG CCGCGCGCG CCGCGCGAG CCGCGCGCG CCGCGCGCG CCGCGCGCG      882

```

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 213 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single

X1: SEQUENCE DESCRIPTION: SEQ ID NO:134:

CCATCAACCA ACCGCTCGCG CCGCCCGCGC CGCCGGATCC GCGTCTGCGG CCACGCCCGC	60
CGGTGCGCTCC GGTGCCCCCG TTGCGCGCGT CCGCGCGGTC GCCCGCGACC GGCTGGGTGC	120
CTAGGGCGGT GTTACCGCCC TGGTTGGCGG GGACGCCGCC GGCACCACCG GTACCGCCGA	180
TGCGCGCGTT GCGCGCGCGG GCACCGTTGC CACCGTTGCG ACCGTTGCCA CCGTTCGCCA	240
CCAGCCACCC GCGCGGACCA CCGGCACCGC CGCGCGCGCC CGCACCGCGG CCGTGCCTGT	300
TGCTGCGCGT ACCCGCGGCA CCGCGGTTGC CCGCGTCACC GCCGACGGAA CTACCGGCGG	360
ACCGCGCGTG CCGCGCGCGG CCGCGCGCAC CGCATTTGGC ACCGCGCTCA CCGCGCGCTG	420
GGAGTCCCGC GATTAGGGCA CTGACCGCGG TAACGAGCGC AAGTACTCTC GGTACCGAG	480
CAGTTCCAGA CCACACCA CAACCGCGTT CTCGCGCGAC TGGGTGAAT GGCAGCCGAT	540
AGCGGCTAGC TGTGCGCTGC GGTCAACCTC CATCATGATC TCGAGGTGAC CGTGACCGCG	600
CGCGCGGAAG GAGGCGGTGA ACTCGCGCTT GAGCGGATCG GCGATCGGTT GGGGCACTGC	660
CCAGGCCAAT ACCGGGATAC CCGGTGTGTA AGCGCGCGCG AGCGAGGCTT CGGTTGCGCG	720
ACGCTGGTGC GCGTGGGCTG TTACCGCGTT CTCGTCGAAC ACGAGTAGCA GGTCTGCTGC	780
CGCGAGGGCA TCCACCACGC GTTGGCTCAG CTCGT	915

2. INFORMATION FOR SEQ ID NO:135:

A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1152 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11. MOLECULE TYPE: DNA genomic

A. SEQUENCE DESCRIPTION: SEQ ID NO:135:

CTAATCGGCT GCGTACGCTC CTGATCAGA CACTCTCGCG AATGACCGCG CGGCTTCAGC	60
CTTTTCCGAG AACAACTGCT GAAGATGCTC CCGCGCGAAA CAGGCGCTCA TTTGACGCTC	120
TATGACGCTC TGAATGAGCA CATGTCGCGC CAGATTCGAT TCGGACCGCT GCGCTAACAC	180
CTATGTAAGA TGTGCACTT CTATGTCGCG CACTGCTGCT CCGCGATCAG TTTGCGCGAG	240
CTATGCTGCA TGTGTCGAG CCAATGCTT CCGGAGAGT AGTATTCGAT TGAAGCTTCT	300

ACACCCGACG TGTCTATCGC GCGCGGGCTC GGTCTAGCAAG TTCACCGCAC CGACGATCCT	480
GCGTTCTGCC TGTCTTAAG CAAGCGGATC GTGTGAGGA AGATCCTGAA TCAGCAGGCC	540
TTGATTGGGG CACACACGTC GGGGCAAGAC GTTGTGAGA GCATCCGCAC GATGAAGCAC	600
TCGCTGGCCT GGGTCGATCG ATCGGGCTCC CTGGCGGAGT TGAACGGGT CGAGGGAAAT	660
GCGCAAAGG CATACTTCAC CGCGCTGGGG CATCTCGTCC CGCAGGAGTT CGCATTCCAG	720
GGCCGCTCGA CTCGGCCGCC GTTGGACGCC TTCAACTCGA TGGTCAGCCT CGGCTATTCC	780
CTGCTGTACA AGAATCATAT AGGGGCGATC GAGCGTCACA GCCTGAACGC GTATATCGGT	840
TTCTTACAC AGGATTCAGG AGGGCAGCGA ACCTCTCTG CGGAATTCCG CACGAGCTCC	900
ACTCAAACG CTGGCCGCTT GTCAGTCCC CTACGTAAT CGGCTGCGCC CAGGCCGCCC	960
CGCCGCGCGA ATACGAGCAG ATCGGACAGC GAATTGCGCC CCAGCCGCTT GGAGCCGTCG	1020
ATACCGGCGG CACTCTACG GCGAGCGAAG AGGCTCGCA CGGTGGCGCC GCGGCTGTCG	1080
CGGTCTACTT CGACACGCGC CATCAGCTAG TCACAGCTCG GCGCGACTTC CATTGCTTGC	1140
TTTGGGCAGC AG	1152

2. INFORMATION FOR SEQ ID NO:136:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 655 base pairs
 - B. TYPE: nucleic acid
 - C. STRANDEDNESS: single
 - D. TOPOLOGY: linear

2. MOLECULE TYPE: DNA genomic

3. SEQUENCE DESCRIPTION: SEQ ID NO:136:

TTTCTTCCGA TTCCGAAAGG TTTACTTATG TTTCTTTTAT TTCCGATGAG TTTGAGGAG	1
TGTAATGCG TAAGAGCAAT GATCTGCT TAAGCAATTT ATGGCTTA TTTGGGCAAT	12
TCTCTGAT CTCCGCTGCG TGGTCTTTT TTCTTAAGG GTATCTGAT TTTCTTCCG	23
TCTGCTTTTC CCGCATTA CCGCGCGGCT GGTTCGCGCG TCCGCTATCG CCGAANGTCC	34
ATCAGCAGAC CCGAGATAGC TTTCTTTGCA AGTTTCTTGA GGTTCGCGCG GGCAGCTTC	45
TCCGCAATT CTATTAAGA GAATTTGCT TATATATTA TCTAGCGAA TTTCTGCTG	56

TCCCGACGGT GGTGCGGGT GCGCCGCGAA AGCGGCGGGT CGGGTGCCAT CAGGAATGCC 540
 TCACCGCCCGC GGCACGTGCAC GGCACGTGCC GCGGCGATGT CAGCCATCGG GACATCATGC 600
 TCGCGTTCAT ACTCCTCGAC CAGTCGGCGC AACAGCTCGA TTCCCGGACC GCCCA 655

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val
 1 5 10 15
 Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Gln
 20 25 30
 Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
 35 40 45
 Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Asp Asp Pro Thr Ala
 50 55 60
 Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asn Val Val Pro Arg
 65 70 75 80
 Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
 85 90 95
 Gln Ala Pro Ala Val Pro Gly Val Val Pro Ala Pro Val Pro Ile Pro
 100 105 110
 Val Pro Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile
 115 120 125 130 135 140 145 150 155 160
 Thr Ile Pro Thr Ala Ile Ile Thr Ile Ile Ile Ile Ile Ile Ile Ile
 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270
 Thr Pro Pro Thr Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr
 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270
 Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr Thr Thr Pro Pro Thr

```

195                200                205
Ala Thr Ala Thr Pro Thr Thr Val Ala Pro Gln Pro Thr Gln Gln Pro
210                215                220
Thr Gln Gln Pro Thr Gln Gln Met Pro Thr Gln Gln Gln Thr Val Ala
225                230                235                240
Pro Gln Thr Val Ala Pro Ala Pro Gln Pro Pro Ser Gly Gly Arg Asn
245                250                255
Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe
260                265

```

2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

A) LENGTH: 174 amino acids
 B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

xii) SEQUENCE DESCRIPTION: SEQ ID NO:138:

```

Ile Asn Gln Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro
1      5      10      15
Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20     25     30
Ser Pro Pro Thr Gly Trp Val Pro Arg Ala Leu Leu Pro Pro Trp Leu
35     40     45
Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50     55     60
Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro Thr
65     70     75     80
Leu Ala Pro Pro Arg Pro Pro Ala Pro Pro Ala Pro Thr Ala Pro Pro
85     90     95
Ala Cys Pro Pro Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser
100    105    110
Pro Pro Thr Leu Leu Pro Ala Asp Ala Ala Cys Pro Pro Ala Pro Pro
115    120    125
Ser Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Pro Ser Ala Ala Pro
130    135    140    145

```

Leu Pro Asp Asp Thr Thr Ala Arg Gly Cys Arg Arg Thr Gly
 165 170

2 INFORMATION FOR SEQ ID NO:139:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Gln Pro Pro Ala Glu Val Ser Asp Gln Arg Val Ser Gly Leu Thr Gly
 1 5 10 15
 Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg
 20 25 30
 Asn Arg Arg
 35

2 INFORMATION FOR SEQ ID NO:140:

- 1 SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 104 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Arg Ala Asp Ser Ala Gly Cys Thr Cys Arg Trp Cys Xaa Pro His Glu
 1 10 15
 Trp Arg Arg Arg Ala Met Arg Ala Gln His Gly Ser Arg Ser Thr Thr
 20 25 30
 Pro Pro Gly Pro Arg His Arg Ser Ala Arg Val Arg Pro Gly Arg Leu
 35 40 45
 Pro Pro Trp Ala Gly Ser Ser Asp Val Phe Pro Pro Trp Phe Ala Ala
 50 55 60
 Leu Met Pro Ala Arg Arg Val Gln Arg Pro Val Trp Pro Xaa Val Asp
 65 70 75

Gly Gln Leu Arg Arg Gln Phe Tyr
100

11 INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1 SEQUENCE DESCRIPTION: SEQ ID NO:141:

GGATCCATAT GGGCCATCAT CATCATCATC ACCTGATCGA CATCATCGGG ACC

53

12 INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR Primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1 SEQUENCE DESCRIPTION: SEQ ID NO:142:

CGTGAATTCG GGGCTGGGTT GGGGCGGGCT CATCTTGAAC GA

42

13 INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR Primer"

(vi) ORIGINAL SOURCE:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T

31

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1. SEQUENCE DESCRIPTION: SEQ ID NO:144:

GGTGAATTC AGCGCTGGAA ATCGTCGCGA T

32

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1. SEQUENCE DESCRIPTION: SEQ ID NO:145:

CGATCCAGCC CTCAGATGAA CACCGATGCC GCT

33

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

[2] INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1993 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

AN ORGANISM: *Mycobacterium tuberculosis*

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 152..1273

(x) SEQUENCE DESCRIPTION: SEQ ID NO:147:

```

TGTTCCTCGA CGGCAGGCTG GTGGAGGAAG GGCCACCGA ACAGCTGTTG TCGTCGCCGA      60
AGCATGCGGA AACGCGCCGA TACGTGCGCG GACTGTGCGG GGACGTCAAG GACGCCAAGC      120
CCGGAATTTG AAGAGCACAG AAAGGTATGG C GTG AAA ATT CGT TTG CAT ACG      180
                               Val Lys Ile Arg Leu His Thr
                               1           5

GTG TTG GCG GTG TTG ACC GGT GCG GCG GTG GTG CTA GCA GCG GCG GCG      240
Leu Leu Ala Val Leu Thr Ala Ala Pro Leu Leu Leu Ala Ala Ala Gly
          10           15           20

TGT GCG TCG AAA CGA CGC ACC GGT TCG GGT GAA ACC GCG GCG GCG GCG      300
Gly Gly Ser Lys Pro Pro Ser Gly Ser Pro Gly Thr Gly Ala Gly Ala
          25           30           35

GGT ACT GTG GCG ACT ACC GCG GCG TCG TCG GCG GTG ACC TTG GCG GAG      360
Gly Thr Val Ala Thr Thr Pro Ala Ser Ser Pro Val Thr Leu Ala Gly
          40           45           50           55

AGC GGT ACC ACC GTG TTG GAG CGC GTG TTG AAG TTG TCG GGT GCG GGT      420
Ser Gly Ser Thr Leu Leu Tyr Trp Leu Thr Asn Leu Thr Gly Pro Ala
          60           65           70           75           80

TTT GAG GAG AGG TAT CGC AAG GTG ACC ATG ACC GGT CAG GCG ACC GGT      480
Phe His Gly Arg Tyr Pro Asn Val Thr Ile Thr Ala Gly Gly Thr Gly
          85           90           95           100

TTT GGT GTC GCG ATT GCG GAG GCG GCG GCG GCG ACC GTC AAG ATT GCG      540
Phe His Ala Gly Ile Asn Ile Ala Ala Ala Gly Thr Val Asn Ile Gly
          105           110           115           120

```

CTG ATG AAC ATC GCG CTA GCC ATC TCC GCT CAG CAG GTC AAC TAC AAC Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Gln Val Asn Tyr Asn 120 125 130 135	556
CTG CCC GGA GTG AGC GAG CAC CTC AAG CTG AAC GGA AAA GTC CTG GCG Leu Pro Gly Val Ser Glu His Leu Lys Leu Asn Gly Lys Val Leu Ala 140 145 150	604
GCC ATG TAC CAG GGC ACC ATC AAA ACC TGG GAC GAC CCG CAG ATC GCT Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala 155 160 165	652
GGC CTC AAC CCC GGC GTG AAC CTG CCC GGC ACC GCG GTA GTT CCC CTG Ala Leu Asn Pro Gly Val Asn Leu Pro Gly Thr Ala Val Val Pro Leu 170 175 180	700
CAC CCC TCC CAC GGG TCC GGT GAC ACC TTC TTG TTC ACC CAG TAC CTC His Arg Ser Asp Gly Ser Gly Asp Thr Phe Leu Phe Thr Gln Tyr Leu 185 190 195	748
TCC AAG CAA GAT CCC GAG GGC TGG GGC AAG TCG CCC GGC TTC GGC ACC Ser Lys Gln Asp Pro Glu Gly Trp Gly Lys Ser Pro Gly Phe Gly Thr 200 205 210 215	796
ACC CTC GAC TTC CCC GCG CTG CCC GGT GCG CTC GGT GAG AAC GGC AAC Thr Val Asp Phe Pro Ala Val Pro Gly Ala Leu Gly Glu Asn Gly Asn 220 225 230	844
GGC CCC ATG CTG ACC GGT TCG GCG GAG ACA CCC GGC TCG GTG GCG TAT Gly Gly Met Val Thr Gly Lys Ala Glu Thr Pro Gly Lys Val Ala Tyr 235 240 245	892
ATG CCC ATG AGC TTC CTC CAC CAG CCC AGT CAA CCC GGA CTC GGC CAG Ile Gly Ile Ser Phe Leu Asp Gln Ala Ser Gln Arg Gly Leu Gly Gln 250 255 260	940
CCC CAA CTA GCG AAT ACC TCT GCG AAT TTC TTC TTC CCC GAC GCG CAA Val Gln Leu Gln Asn Ser Ser Gly Asn Pro Leu Leu Pro Asp Ala Gln 265 270 275	988
ATG ATT CAG GCG TTC GCG GGT GGT CTC GCG CAA ACC CTC GCG AAC Ser Ile Gln Ala Val Val Ala Gly Phe Ala Ser Lys Thr Pro Ala Asn 280 285 290 295	1036
CAG GCG ATT TTC ATG ATG CAC GCG GCG GCG GCG CAG GCG TAC CCG ATC Gln Ala Ile Ser Met Ile Asp Gly Pro Ala Pro Asp Gly Tyr Pro Ile 300 305 310	1084
ATG AAT TAC CAG CAG GAT ATG GT AAT AAT GGT CAA AAG CAG GCG GGT Met Asn Tyr Cag Cag Gat Atg Gt Aat Aat Ggt Caa Aag Cag Gcg Ggt 315 320 325	1132

50 55 60
 Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr
 65 70 75 80
 Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
 85 90 95
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly
 100 105 110
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
 115 120 125
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
 130 135 140
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 145 150 155 160
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220
 Ala Leu Gln Gln Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Gln
 225 230 235 240
 Thr Pro Gly Lys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255
 Ser Gln Arg Gly Leu Gly Gly Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270
 Thr Ser Leu Thr Asn Ala Gln Ser Ile Gln Ala Ala Ala Gly Ile
 275 280 285
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
 290 295 300
 Ala Ser Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
 305 310 315 320
 Asn Arg Gln Thr Asp Ala Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
 355 360 365

Ile Ala Thr Ile Ser Ser
 370

(2) INFORMATION FOR SEQ ID NO:149:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1993 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TGTTTCTTCCA	CGGCAGGCTG	GTGGAGGAAG	GGCCCAACCGA	ACAGCTGTTG	TGCTCGCCGA	50
AGCATGCGGA	AACCGCCCCA	TACGTGCGCG	GACTGTGCGG	GGACGTCAAG	GACGCCAAGC	120
GCGGAAATTC	AAGAGCACAG	AAAGGTATGG	CGTGAAAAAT	CGTTTGCATA	CGGTGTGGGC	180
CGTGTGACG	GCTGCGCCGC	TGCTGCTAGC	AGCGGCGGGC	TGTGGCTCGA	AACCACCGAG	240
CGGTTGCGCT	GAAACGGGCG	CGGGCGCCCG	TACTGTGCGG	ACTACCCCGG	CGTGTCGCCC	300
GGTGACCTTC	GGGAGACCG	GTAGTACGCT	GCTGTACCGG	GTGTTCAACG	TGTGGGGTCC	360
GCGGTTTCAC	GAGAGGTATC	CGAAGCTCAC	GATCACCGCT	CAGGGCACCG	GTTCTGGTGC	420
GCGGATCGCG	GAGCGCGCGG	GCGGACGGST	GAAGNTTGGG	GCTTCGGACG	GCTATCTGTC	480
GGAAGTGCAT	ATGCGCGCGG	AGAAAGGGCT	GATGAACATC	TGCTAGCGCA	TGTCCGCTCA	540
GAGGCTGAAG	TAGAACCTGC	GCGGAGTCAG	GAAGGACCTG	AAGCTGAACG	GAAGAAGTCT	600
GCGGCGCATG	TACGAGGGCA	GCATCAAAAC	GTGGGACGAC	CGGCAGATCG	GTGCGCTCAA	660
GCGGCTTCTG	GAGGTCGCGG	GAGGCGGGCT	AGTTGCTGCT	TAGGCTGCGG	AGGCTGCGGG	720
GAGGCTTCTG	GTGTCGCGG	AGTATCTGCT	GAAGGAGATC	AGTGAAGGCT	GAGGCAAGTC	780
GAGGCTTCTG	GTGTCGCGG	AGTATCTGCT	GAAGGAGATC	AGTGAAGGCT	GAGGCAAGTC	840
GAGGCTTCTG	GTGTCGCGG	AGTATCTGCT	GAAGGAGATC	AGTGAAGGCT	GAGGCAAGTC	900
GAGGCTTCTG	GTGTCGCGG	AGTATCTGCT	GAAGGAGATC	AGTGAAGGCT	GAGGCAAGTC	960
GAGGCTTCTG	GTGTCGCGG	AGTATCTGCT	GAAGGAGATC	AGTGAAGGCT	GAGGCAAGTC	1020

```

GACGTTTCAG GCATTTCTGC ACTGGGCGAT CACCGACGGC AACAGGGCTT GTTCTCTCGA      1200
CCAGGTTTCAT TTCCAGCCGC TCCCCCCCGC GGTGGTGAAG TTGTCTGACG CGTTGATCGC      1260
GACGATTTCC AGCTAGCCTC GTTGACCACC ACGCGACAGC AACCTCCGTC GGGCCATCGG      1320
GCTGCTTTGC GGAGCATGCT GCGCCGTGCC GGTGAAGTCG GCCGCGCTGG CCGGGCCATC      1380
CGGTGGTTGG GTGGGATAGG TGCGGTGATC CCGCTGCTTG CGCTGGTCTT GGTGCTGGTG      1440
GTGCTGGTCA TCGAGGCGAT GGGTCCGATC AGGCTCAACG GGTTCGATTT CTTCACCGCC      1500
ACCGAATGGA ATCCAGGCAA CACCTACGGC GAAACCGTTG TCACCGACCC GTGGCCCATC      1560
CGGTCCGGCC CTACTACGGG GCCTTCCCGC TGATCCTCGG GACGCTGGCG ACCTCGGCAA      1620
TCCCCCTGAT CTTCCCGGTC CCGGTCTCTC TAGGAGCCGC GTGGGTGATC GTGGAACGGC      1680
TCCCGAAACG GTTGGCCGAG GCTGTGGGAA TAGTCTCGGA ATTGCTCCGC GGAATCCCCA      1740
CCCTGCTGCT CGGTTTGTGG GGGGCAATGA CTTTCGGGCC GTTCATCGCT CATCACATCG      1800
CTCCGCTGAT CGCTCACAAC GCTCCCGATG TGCCCGTGCT GAACTACTTG CCGGGCGACC      1860
CGGCAACCGG CGAGGGCATG TTGCTCTCCG GTCTGTGTTT GCGCGTGATG GTCTTTCCCA      1920
TTATCGCCAC CACCACTCAT GACCTGTTCG GGCAGGTGCC GGTGTTGCCC CCGGAGGGCG      1980
GATCGGGAA TTC

```

1993

2. INFORMATION FOR SEQ ID NO.150:

A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 374 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

3. SEQUENCE DESCRIPTION: SEQ ID NO.150:

```

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
10
Glu Leu Leu Ala Ala Ala Gly Tyr Gly Ser Lys Leu Thr Ser Gly Ser
20
Phe His Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
30
Met Phe Thr Thr Thr Ala His Thr Gly Ser Thr Leu Leu Tyr Pro Leu
40

```

Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
 85 90 95
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly
 100 105 110
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
 115 120 125
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
 130 135 140
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 145 150 155 160
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220
 Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 225 230 235 240
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 275 280 285
 Ala Ser Val Thr Tyr Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Ser
 290 295 300
 Ala Pro Arg Gly Tyr Arg Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
 305 310 315 320
 Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
 325 330 335
 His Thr Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asn Gln Val
 340 345 350

370

(2) INFORMATION FOR SEQ ID NO:151:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1777 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

```

GGTCTTGACC ACCACCTGGG TGTGGAAGTC GGTGCCCCGA TTGAAGTCCA GGTACTCGTG      60
GGTGGGGCGG GCGAAACAAT AGCGACAAGC ATGCGAGCAG CCGCGGTAGC CTTTGACGGT      120
GTAGCGAAAC GGCAACGCGG CCGCGTTGGG CACCTTTTTC AGCGCTGATT TGCACAACAC      180
CTCGTGGGAG GTGATGCGGT CGAATTGTGG CCGCGCAACG CTGCGGACCA GCGCGATCGG      240
CTGCAACCGG GCAGCGCGCG TCGTCAACGG GCATCCCGTT CACCGCGACG GCTTGCCCGG      300
CGCAACGCAT ACCATTATTC GAACAACCGT TCTATACTTT GTCAACGCTG GCGCGTACCG      360
AGCGCGCGAC AGGATGTGAT ATGCCATCTC TGCGCGCACA GACAGGAGCC AGGCCTTATG      420
ACAGCATTCG GCGTCGAGCC CTACGGGCAG CGGAAGTACC TAGAAATCGC CCGGAAGTGC      480
ATGCGGTATA TCGACGAAGG CAAGGGTGAC GCGATCGTCT TTCAGCACGG CAACCCGACG      540
TGGTCTTACT TGTGGCGCAA GATGATGCGG CACTTGBAAG GGCTGGGCGG GGTGGTGGCG      600
TGGGATCTGA TGGGATGGG CGCCTCGGAC AGCTCAGCG CATTGGGACC CGACCGCTAT      660
AGCTATGCGG AGCAACGAGA CTTTCTCTTC GCGCTGTGGG ATGCGCTCGA GTTCGGGCAU      720
GACGTGGTAC TGGTGTGSCA CGAATGGGGG TCGGCGCTCG GCTTCGACTG GCTTAACGAG      780
GATCGCGAAG GATGCGAGGG GATCGCGTTC ATGGAAGCGA TCGTCACCGG GATGACGTGG      840
TGGGATGCGG AGCGCGCTGT TCGGCTGTTC TTGGAAGCTT TGGGATGCGG TGAAGGCGAG      900
TGAATGCGCT TGAAGCGAAG GATCTTGTTC GAAGGCTGCT TGAAGGCGCT GATCTTGTGA      960
GAGTGTGCGG AGGAGGAAT GAAGGATAT CCGCGGAGAT TGTGGAAGG TGGGAGGAT      1020
CTTGGCGTGA GCTTGTGTTC TCGAGGAAGT CTTGGAATTC AGGCTGAGCG GCGCGAGGTC      1080
GTGCTGTTCG TGAAGGAGTA TCGGAGGTGG CTGAGGAGAA TCGACATGCT GAACTCTTTC      1140
ATGAAGGCGG AGCGCGCGCG GATGATGAGT GGTGCGATTC GTGATGATTC GAGGAGCTGG      1200

```

GACCAAGAAT GTGATTTCCG GCGAAGGCGG CGCCCTGCTT GTCAACTCAT AAGACTTCCT 1380
 GCTCCGGGCA GAGATTCTCA GGGAAAAGGG CACCAATCGC AGCCGCTTCC TTGCAACGA 1440
 GGTGACAAA TATACGTGGC AGGACAAAGG TCTTCCTATT TGCCGAGCGA ATTAGTCGCT 1500
 GCCTTTCTAT GGGCTCAGTT CGAGGAAGCC GAGCGGATCA CGCGTATCCG ATTGGACCTA 1560
 TTGAACCGGT ATCATGAAAG CTTTGAATCA TTGGAACAGC GGGGGCTCCT GCGCCGTCGG 1620
 ATCATCCCAC AGGGCTGCTC TCACAACGCC CACATGTACT ACCTGTTACT AGCGCCGAGC 1680
 GCCGATCGGG AGGAGGTGCT GCGCGGTCTG ACGAGCGAAG GTATAGGCCG GGTCTTTTAT 1740
 TACGTGCGGC TTCACGATTC GCGGGCCGGG CGTCGCT 1777

2. INFORMATION FOR SEQ ID NO:152:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGATTGAAT CGTACCGGTC TGCTTAGCGG CTGCGTCCCG TGAATGCCCA TATCAGGCAC 50
 GCGATGTTTC TGGGTGTCCA CTTTCGCCCC ATGCCCGGAC GTTGGTAAAC CCAGGGTTTG 100
 ATGATTAATT CCGGGGGAGC TTTCGGGAA GCGGGCCAGG ATGTGGTGA GCGGGGGGCT 150
 TTTTGTGGG TATGCGACCG CTGATGCTC AGCCCCGCTG GCGCGAGCTA GCGAGGTTTT 200
 TTTTGTGTC GTGACAGTC GTACTCCGCT GACGACCCCG GCGGGTGCCT GGTGAAGGAC 250
 CTTGACCGAC CCGGGCGATT CAGA 324

INFORMATION FOR SEQ ID NO:153

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 128 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:153

CAATAACG CCGCTTCCG TGGCAAGGCA CTTGTACCA GTCAACCACT TCGCCTCCCG 50

```

GCACGACTTC CAGCCCGACT CGATCGGCGT GGTGACCCGT CCTGTCCCTA TGGCTGCCTG      240
GGAAGCTCGC GTTCGGAAGC GATTTCGGTT CCTCACTGAC CTCGACGCCG ACGAGCAGCG      300
GTGGGCCCGC TGGACGAAC GGCACCGCCG CGAAGTGGAG AACCGCTGG CGGTGCTGCG      360
GTCTGATCA ACCTGCCGGC GATCGTGCCG TTCCGCTGGC ACGGTTGCGG CTGGACGCGG      420
CTGAATCGAC TAGATGAGAG CAGTTGGGCA CGAATCCGGC TGTGCTGGTG AGCAAGACAC      480
GAGTACTGTC ATCACTATTG GATGCACTGG ATGACCGGCC TGATTCAGCA GGACCAATGG      540
AATGCCCCG GCGAAAACGT CTCGGAGATG ATCGGCGTCC CCTCGGAACC CTGCGGTGCT      600
GGCTTCATTG GGACATCGGT CCGGCTCGCG CGATCGTGGT GACGCGAGCG CTGAAGGAGT      660
GGACGCGCGG GGTGGACCGG TCGTGGAC CCGGCGAGAG GGTGCTGCTG CCGAAGCGCG      720
GGATCGCGCA GAAGCGCTTC GAGGTGCGCG CGGAGGAGTT GTTGTGTTTC CCGACGCTCG      780
CGCACAGCCA CGCGGAGCGG GTTCGCGCGG AGCAGCGCGA CCGTCTGCGG CCGGCGCGCG      840
CGCACAGCAC CGACGAGTGT GTGCTACTGG GCGCGCGAGC GAAAGTTGTT GCGGCACTGC      900
CGGTTAACC GCGAGAGGGT CTGGACGCGA TCGAGGATGT GCACATCTGG ACCGCGGAGT      960
CGGTGCGCGC CGACCGGCTC GACTTTGCGC CGAAGCAGAA ACTGGCGCTC TTGGTGGTGT      1020
CGGCGATCCC GGTGCGCGAG CCGGTGCGCG TGGCGCGTAG GCGCGAGTAG GCGGCTGCA      1080
CGAGCTGGGT CGAGCTGCGG GTGACGCGCA CCGTGGCGCG GCGGCTGCGG GCGAGGCGCG      1140
CGGTGCGCGA GGTGCGCGCG CGGTGCGCGG AGCGGCTGCG TTGAGTGGCG GCGATCGCTT      1200
CGGTGCGAGC TGTAGCGCGA GTGCGCGCTG CGAATGATGT CCGTGGCGTT CGGTGCGCTG      1260
CGGCTGCAAT TGAAGCGCGG CGCACAGCA CCGTGGCGCG GCGATGCTG CCGGCGCGCG      1320
CGGCGCGCGG GTAGAGCG

```

1. INFORMATION FOR SEQ. NO. 154

(A) SEQUENCE CHARACTERISTICS

- (1) LENGTH: 151 base pairs
- (2) TYPE: nucleic acid
- (3) STRANDEDNESS: single
- (4) TOPOLOGY: linear

(B) SEQUENCE DESCRIPTION: SEQ. NO. 154

CGGCGCGCGG GTAGAGCG

```

GCACCGGCGG TCGCGGCGGC GCCGGCAACG ACGCGGCGAG CACCGGCAAT CCGGCGGGTA      240
AGGGCGGGCGA CGGCGGGATC GCGGTGCGG CCGGGGCGCG GGGCGCGGCC GGCACCGGCA      300
ACGGCGGGCCA TCGCGGCAAC C                                          321

```

(2) INFORMATION FOR SEQ ID NO:155:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 492 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1: SEQUENCE DESCRIPTION: SEQ ID NO:155:

```

GAAGACCGCG CCGCGGCGATA TCGATCGGCT GCGCGACTAG TTTCGCGGAA CGTGACCGCG      60
CGCGCGTTCG GCTGATGATC ACGGCTGGCT ACGCGCGGAA CCGCACCGGA TGGGTGCTGC      120
CGTTGCGCTC CGAACTCGTC AATTGCGGCG AAGCGGAGCG GCACGCGGGA ATGACGAGCG      180
CGGTCCACGA TTGCGGTGCA AAGATCGTGC TGCAAATCGT GCACGCGGGA TGTACGCTT      240
ATCACCGCACT TCGGCTCAGC GCGTCGCGGA TCAAGGCGCG GATCACCGCG TTTCGTGCGC      300
TACCACTATC GCGTCGCGGG CTCGAAGCGA GCATCGCGGA TTTCGCGGCG TCGGCGCAGT      360
TCCCCGCGGA TCGCGGCTAC GACGGCGTCC AAATCATGCG CAGCGAAGGG TATCTGCTCA      420
ATCACTTCCT GCGCGCGGCG ACCAACAAGC GCACCGAATC CTGGGCGCGC ACACCGGCGA      480
ATCTTGGCGG GT                                          492

```

2. INFORMATION FOR SEQ ID NO:156

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 536 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS
- D. TOPOLOGY: linear

X1: SEQUENCE DESCRIPTION: SEQ ID NO:156

```

Pro Ala Glu His Leu Val Met Ile Asp Ala Val Gln Leu Tyr Arg Ala
1          5          10          15
Asp Ala Ala Asp Gln Ala Asp Pro Leu Glu Pro Glu Ser Ala Arg Arg
16          20          25          30          35          40

```

50	55	60
Ser Ala Ala Gln Asp Val Ile Cys His Leu Cys Pro His Arg Gln Glu		
65	70	75 80
Pro Gly Leu Met Thr Ala Phe Gly Val Glu Pro Tyr Gly Gln Pro Lys		
	85	90 95
Tyr Leu Glu Ile Ala Gly Lys Arg Met Ala Tyr Ile Asp Glu Gly Lys		
	100	105 110
Gly Asp Ala Ile Val Phe Gln His Gly Asn Pro Thr Ser Ser Tyr Leu		
	115	120 125
Trp Arg Asn Ile Met Pro His Leu Glu Gly Leu Gly Arg Leu Val Ala		
	130	135 140
Cys Asp Leu Ile Gly Met Gly Ala Ser Asp Lys Leu Ser Pro Ser Gly		
	145	150 155 160
Pro Asp Arg Tyr Ser Tyr Gly Glu Gln Arg Asp Phe Leu Phe Ala Leu		
	165	170 175
Trp Asp Ala Leu Asp Leu Gly Asp His Val Val Leu Val Leu His Asp		
	180	185 190
Trp Gly Ser Ala Leu Gly Phe Asp Trp Ala Asn Gln His Arg Asp Arg		
	195	200 205
Val Gln Gly Ile Ala Phe Met Glu Ala Ile Val Thr Pro Met Thr Trp		
	210	215 220
Ala Asp Trp Pro Pro Ala Val Arg His Val Phe Gln Gly Phe Arg Ser		
	225	230 235 240
Pro Gln Gly Glu Pro Met Ala Leu Glu His Asn Ile Phe Val His Arg		
	245	250 255
Val Leu Pro Gly Ala Ile Leu Arg Gln Leu Ser Asp His Glu Met Asn		
	260	265 270
His Tyr Arg Arg Pro Phe Val Asn His His Glu Asp Arg Ala His Thr		
	275	280 285
Leu Ser Trp Pro Asp Asn Leu Pro His Asp Gly Glu Pro Ala His Val		
	290	295 300
Val Ala Leu Val Asn His Tyr Arg Ser Trp Leu Glu Glu Thr Asp Met		
	305	310 315 320
Trp Leu Leu Phe Thr Arg Ala Gly His Thr Thr Thr Thr Thr Thr Thr		

Pro Gly Val His Phe Val Gln Glu Asp Ser Asp Gly Val Val Ser Trp
 355 360 365
 Ala Gly Ala Arg Gln His Arg Arg Pro Gly Ser Ala Leu Ile Ser Arg
 370 375 380
 Asp Gln Glu Cys Asp Phe Arg Arg Arg Arg Arg Pro Ala Cys Gln Leu
 385 390 395 400
 Ile Arg Leu Pro Ala Pro Gly Arg Asp Ser Gln Gly Lys Gly His Gln
 405 410 415
 Ser Gln Pro Leu Pro Ser Gln Arg Gly Arg Gln Ile Tyr Val Ala Gly
 420 425 430
 Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
 435 440 445
 Ala Gln Phe Glu Glu Ala Glu Arg Ile Thr Arg Ile Arg Leu Asp Leu
 450 455 460
 Trp Asn Arg Tyr His Glu Ser Phe Glu Ser Leu Glu Gln Arg Gly Leu
 465 470 475 480
 Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met
 485 490 495
 Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Glu Glu Val Leu Ala
 500 505 510
 Arg Leu Thr Ser Glu Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu
 515 520 525
 His Asp Ser Pro Ala Gly Arg Arg
 530 535

2. INFORMATION FOR SEQ ID NO:157

SEQUENCE CHARACTERISTICS

- A LENGTH: 534 amino acids
- B TYPE: amino acid
- C STRANDEDNESS:
- D TOPOLOGY: linear

3.1 SEQUENCE DESCRIPTION SEQ ID NO:157

Asp Glu Ser Ala Pro Arg Ser Pro Met Leu Glu Ser Ala Arg Pro Arg
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

Gly Asp Asp Arg Ala Gly Leu Gly Val Asp Glu Gln Phe Arg His Val
 50 55 60
 Gly Phe Leu Glu Pro Ala Pro Val Leu Val Asp Gln Arg Asp Asp Leu
 65 70 75 80
 Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly
 85 90 95
 Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His
 100 105 110
 Phe Leu Val Ala Glu Leu Ser Gln Asp Arg Pro Gly Gln His Pro Phe
 115 120 125
 Asp Lys Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser
 130 135 140
 Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg
 145 150 155 160
 His Arg Gly Asp Asp Arg Phe His Glu Arg Asp Pro Leu His Ser Val
 165 170 175
 Ala Met Leu Val Ser Pro Val Glu Ala Glu Arg Arg Ala Pro Val Val
 180 185 190
 Gln His Gln Tyr His Val Val Ala Glu Val Glu Arg Ile Pro Glu Arg
 195 200 205
 Gln Gln Lys Val Ser Leu Leu Ala Ile Ala Ile Ala Val Gly Ser Arg
 210 215 220
 Trp Ala Glu Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His
 225 230 235 240
 Gln Pro Ala Gln Pro Phe Gln Val Arg His Asp Val Ala Pro Gln Val
 245 250 255
 Arg Arg Arg Gly Val Ala Val Leu Leu Ser Ser Gly Val Thr Leu Ala
 260 265 270
 Ser Val Asp Ile Arg His Ala Leu Ser Gly Asp Phe
 275 280

1 INFORMATION FOR SEQ ID NO:158

1 SEQUENCE CHARACTERISTICS
 A LENGTH 280 base pairs
 B TYPE DNA

ATGAACATGT CGTCGGTGGT GGGTCGCAAG GCSTTTGCGC GATTCGCCCG JTACTGCTCG	60
GGCATGCACG CGATCGCCCG TTTCTCCGAT GCSTTGCSCC AAGAGCTGCC GGGTAGCCGA	120
ATCCCGCTCT CGGTGATCCA CCGGGCCTG ACCGAGACAC CGCTGTTGGC CAACGTCGAC	180
CCCGCCGACA TGCCGCCCGC GTTTCGCAGC CTCACGCCCA TTCCCGTTCA CTGGGTCCGC	240
GCAGCGGTGC TTGACGGTGT GGCG	264

2. INFORMATION FOR SEQ ID NO:159:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TAGTGGGGA CGATGACCTT GCGGTCCAGG CCGACGCTT CAAGCACCAG CCGGACCAG	60
AAGCGCTTC GATCCTTACC CCGGAAGCAG TGGGTGAGCA CCGGGCGTCC GCGCGCAATC	120
AGTGTGAGG CACGATGTAG CCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGCGGATAC	180
TTCTGCTCA TGTAGCGGCT GCGCGCTCA TTTATCGACT GGCTGGATTC CCGGGACTTC	240
CGCTTGAGG CGTCACTGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	300
TTCTGCTCA TGTAGCGGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	360
AGTGTGAGG CACGATGTAG CCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGCGGATAC	420
TTCTGCTCA TGTAGCGGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	480
CGCTTGAGG CGTCACTGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	540
AGTGTGAGG CACGATGTAG CCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGCGGATAC	600
TTCTGCTCA TGTAGCGGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	660
AGTGTGAGG CACGATGTAG CCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGCGGATAC	720
TTCTGCTCA TGTAGCGGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	780
AGTGTGAGG CACGATGTAG CCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGCGGATAC	840
TTCTGCTCA TGTAGCGGCT TACGAGCTTC TTGAATGCGG TTTCGTCCGG CGCTGAGTCC	900

CGCACCACCG TCGGTTCCGA CGTACGGACC GGGTCCGACA CCATGTTCTT GGCCCCAGTA 1080
 ACCATCGGCG ACGGCGCGTA TACCGGGGCC GGCACAGTGG TCGGGGAGGA TGTCCCGCCG 1140
 GGGGCGCTGG CAGTGTGGGC GGGTCCGCAA C 1171

(2) INFORMATION FOR SEQ ID NO:160:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:160:

JCAAAGGCGG CACCGGGGGG GCGGCGATGA ACAGCCTCGA CCGGCTGCTA GCGCCCCAAG 60
 AGGCGGGGCA AGGCGGCGAC GCGGCGACCG GCGGCAACCG GCGGCGGGCG GCGACCAAGT 120
 TCACCAAGG GCGGCGGCGG AACGCGGGCA ACGGCGGTGA GCGGCGGGTC GCGGCGAAGC 180
 GCGGAAACGG GCGGAAACGG GCAGACAACA CCACCAACCG GCGGCGG 227

(2) INFORMATION FOR SEQ ID NO:161:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 304 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:161:

CTGCGGCGG ATGCGGCGGG AGGCGGCTAG CGTGGGCGG GGTGTATGCT CAGGCGGCA 60
 GCGGCGGCGG GGTTCAGCTC CAACCAAGCG GCGGCGGCG GCGGCGGCG GCAACGCGCG 120
 TACTGCGCA CTGCTCGGCG GCAACGCGCG TCACGCGCG AATGCGGCGA ACGGCGGCG 180
 TCTGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG 240
 TCTGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG 304

(2) INFORMATION FOR SEQ ID NO:162:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 base pairs
- (B) TYPE: nucleic acid

GTGGGACGCT GCCGAGGCTG TATAACAAGG ACAACATCGA CCAGCGCGCG GTGGGTGAGC 60
 TGATCGACCT ATTTAACAGT GCSCGCTTCA GCGGGCAGGG CGAGCACCGC GCGCGGGATC 120
 TGATGGGTGA GGTCTACGAA TACTTCCTCG GCAATTTTCG TCGCGCGGAA GGAAGCGGG 180
 GTGGCGAGTT CTTTACCCCG CCCAGCGTCG TCAAGGTGAT CGTGGAGGTG CTGGAGCCGT 240
 CGAGTGGGCG GGTGTATGAC CCGTGCTGCG GTTCGGGAGG CATGTTTGTG CAGACCGAGA 300
 AGTTCATCTA CGAACACGAC GCGGATCCGA AGGATGTCTC GATCTATGGC CAGGAAAGCA 360
 TTGAGGAGAC CTGGCGGATG GCGAAGATGA ACCTCGCCAT CCACGGCATC GACAACAAGG 420
 GGTGCGGCGC CGGATGGAGT GATACCTTCG CCGCGACCA CCACCCCGAC GTGCAGATGG 480
 ACTACGTGAT GCGCAATCCG CGCTTCAACA TCAAGACTG CCGCGCGAAC GAGGAAGACC 540
 CACGCTGCGC GTTGGGTGTT GCGCGCGCGA ATAACGCCAA CTACCCATGG ATTGAGCACA 600
 TGCTGTACAA GTTGGCGCGG GAGGCTCGG GCGCGGTGGT GATGCGCAAC GCGTCGATGT 660
 CGTCGAACTC CAACCGCAAG GGGGATATTC GCGCGCAAAAT CGTGGAGGCG GATTTGCTTT 720
 CGTCGATGCT CGCGTTACCG ACCCAGCTGT TCGCGAGCAC CGGAATGCGG GTGTGCTGT 780
 CGTTTTTCGG CAAAAACAAG GCGGCAGGTA AGCAAGGCTC TATCAACCGG TCGGGGCGAG 840
 TGCTGTGAT CGACGCTGCT GAAGTGGCGC AGTAGTGA CCGCGCGGAG GCGCGGTGA 900
 CGAACGCGA GATGCTCGCG ATCGGGGATA GTTCGACCG GAGTACGAGG AGCGCGAAG 960
 CGCGCTCGCG TGCTGCGCGG GGTAACTCGG GAGTGGGCT TAACTGCGG GCGGTGCTG 1020
 CGCGCGCGCG CGGCAACCGG GGTGTGCGCG CGCTGTGCTT CGGCAACGCT GTGGGCGCGG 1080
 AGCGCGCGCA CGCGCGCAAC CGCGGCGAAG CGCGCGACCG CAGGAGGCGG GCGCGCGCGG 1140
 TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG 1200
 TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG 1260
 TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG 1320
 TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG 1380
 TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG TAACTGCTG 1440

(D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GGGCGGCGG GGGCGGATTT TCTCGTGCTT TGATTGTGCG TGGGGATAAC GCGCGTGATG	60
GTGGTAACGG CGGGATGGGC GGGGCTGGCG GGGCTGGCGG CCGCGGCGGG GCGCGCGGCC	120
TGATCAGCCT GCTGGGCGGC CAAGGCGCGG GCGGGGCGCG CGGGACCGGC GGGGCGGCG	180
GTGTTGGCGG TGACGGCGGG GCGGGGCGGC CCGGCAACCA GGCCTTCAAC GCAGGTGCCG	240
GCGGGGCGCG CGGCTGATC AGCCTGCTGG GCGGCCAAGG CCGCGGCGGG GCGGCGGGA	300
GCGGCGGCGG CGGCGGTGTT GCGGTGAC	329

2) INFORMATION FOR SEQ ID NO:164:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:164:

GCAACGGTGG CAACGGCGCG ACCAGCAGCA CGGTGGGGAT GCGCGGAGGT AACTGTGGTG	60
GCGCGGGGCT GATCGGCAAC	80

2) INFORMATION FOR SEQ ID NO:165:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 192 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:165:

TTTCTTCTTC TACTGAGG GAGGAGTTC ATGAGGCTTC GCGGCGAAT ATGAGGCTTA	60
AGGCTTATTA GTTATCTTC GAGGAGGCTT ATGCAAGGT TGGGCTGAGT GCGGAGGAA	120
TAAAGTATAT TGAGGCGGAG GCGGAGGAG GCGGCTGCTG GCGGCTCGG GCAGGATCGG	180
TTGCTTCTCA TTGCGGCGG CAAGGCGGCT GATGCTCGG AACGGCGGCG ATCTGTGAG	240
TAAATATAT TATATTAAT GAGGAGGCTT AATGCTCAA GCGGCTTCTG AAGGAGGGA	300
TTTCTTCTTC TACTGAGG GAGGAGTTC ATGAGGCTTC GCGGCGAAT ATGAGGCTTA	360

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 535 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

```

ACCGGCGCCA CCGGCGGCAC CGGGTTCGCC GGTGGCGCCG GCGGGGCCCC CGGGCAGGGC      60
GGTATCAGCG GTGCGGCGCG CACCAACGGC TCTGGTGGCG CTGGCGGCAC CGGCGGACAA      120
GGGCGGCGCG GGGGCGGTGG CGGGGCGCGC GCGGATAACC CCACCGGCAT CGGCGGCGCC      180
GGGCGCACCG GCGGCACCGG CGGAGCGGCG CGAGCGGGCG GGGCGGGTGG CCGCATCGGT      240
ACCGGCGCCA CCGGCGGCAC GGTGGCGAGC GTGCGTAACG CGGGGATCGC CGGTACCGCG      300
GGTACCGGTC GTGCGGTGGG TGCTGGTGGT GCAGGTGCGG CTGCGGCGCG TGGCAGCAGC      360
CGTACCGGTC GCGCGCGGTC CGGCGGCGCG GTGCGCGGAG AAGCGCGAGC GCGCGGCAAC      420
AGCGGTGTGG GCGGCACCAA CGGCTCGGCG GCGCGCGGCG GTGCAGGCGG CAAGGGCGGG      480
ACCGGAGGTG CCGGCGCGTC CGGCGCGGAC AACCGCACCG GTGCTGGTTT CGCGG      535

```

2. INFORMATION FOR SEQ ID NO:167:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 590 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:167:

```

CGAAGCTCCG CGGGGCGGATA CGGGGCTCAC CGACTACTAC ATGATCCGCA CGGAGAATCG      60
GAGGCTGCTG GAAGCGCTTC CGGGGCTGCT GTTATCTGGA GATTTGCTGG CGGAGCTGAT      120
GATTTGGAAG TGAAGCTGA TTCTCAAGCT AGGTACGGT GATTTGAANT AAGGCTACTT      180
GATTTGCTGA CGGATGTGTC GAAGCGCTTC TCGGCTGCTG CGGAAGCTTC CGGCTCAGGT      240
GATTTGCTGAT GGTGCTGCTG CGGGAACAGA AGAAGCGATC GTTCACTTCA CGGCGGACCT      300
GATTTGCTGTC CGGCTGGAAG GGTTCAGGCT CGGCGGATC CAGCTGCGCG AAGCGGCGGA      360
GATTTGCTGTC CGGCTGGAAG GGTTCAGGCT CGGCGGATC CAGCTGCGCG AAGCGGCGGA      420

```

ATCAACGGCA TCGGCTATCC CCGGGCGGCC ACCCTAGGTT TAGGCACGAT CGATAGCGGG 500
 CGGGCTGGAA TTGCTCAGCC TCCTCGCGGC GGCCTCGGAC ACCGTTGAA ACATCGAGGG 560
 CCGCTCACC TAACGGATTC CCGACGGCAT 590

(2) INFORMATION FOR SEQ ID NO:168:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:168:

ACGGTCACGG CCGTACTGGC GCGGGCCAGC GCGGCAACGG CCGGAATCCC GGGTGGCTCT 60
 TGGGCACAGC GGGGGGTGGC GCGAACGGTC GCGGCGCCAG CACCGGTACT GCAGGTGGCC 120
 GGTCTGGGGG CACCGGCGGC GACGGCGGGA CCGGCGGGCC TGGCGCCCTG TTAATGGGGC 180
 CCGGCGGGCG CCGGCACGGT GCGACTGGCC GCGGCGGGCG TGGCGGTCTC GACCGTGGCC 240
 GCGGCGGGCG GCGGCGGGCG GCGGCGGGCA ACAGCGGGCG CCGGGGTCAA GCGGCGCTGC 300
 TGTTCGGGGC CCGGCGGACC GCGGAGCCG GCGGCTACGG CCGGCTATGG GGTGGCGGGC 360
 GTGACGGCTT CGACGCGACC ATGGCGGGCT TGGTGGTAC GGTGGC 407

(2) INFORMATION FOR SEQ ID NO:169:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 468 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:169:

ATGCTTAAI CCGATCAGC GCGGGCGAA GCGATTGCGC GTCTTACCA CAGAACATCG 60
 TGAATGCGC GCGGCGGACC CCGGCGCTCG GTCTGCGGCG GTGAAAGCGC TGAATGAGC 120
 ACAGCGAGTC GTTGGCGGGC TCGGAGGCGA ACAGCTCGGT GTGACCGGTG TAGATCGGCG 180
 GATGCGGGC GTTGGCGGAC GCGATTGCGC ACAGCGGGCG GTCTTGTGA TCGTGGACGA 240
 TTAGGCGAT GTCTGCGGCG ACAGCGGGCG GCGGCGGAA GGTGGCGGCG GTGCGGAGTA 300
 GCGGCGGAA GTCTGCGGCG ACAGCTGCGC GCGATTGCGC GCGGCGGCG GTGCGGAGTA 360

12. INFORMATION FOR SEQ ID NO:170:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

```

GGTGGTAACG GCGGCCAGGG TGGCATCGGC GCGCCCGCG AGAGAGCCGC CGACGGCGCC      60
GGCCCCAATG CTAACGGCCG AAACGGCGAG AACGGCGGTA GCGGTGGTAA CCGTGGCGAC      120
GGCGCGCGCG GCGGCAATGG CGGCGCGCGG GCAACCCGC AGGCGGCGCG GTACCCGAC      180
GGCGCCACGG GCACCGCGCG GCACCGCGCG AACGGCGGC      219

```

1. INFORMATION FOR SEQ ID NO:171:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 494 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

```

TAGCTCCCGC GAGGCGCGCA AGGCGCGGCA GCGTGGCCAC GGCGGTGACG GCGTCCGCGC      60
TACAGTTCC CTACCCGAG GCGGCGCGG GCGTGGCGCG GCGCCCGCGG GCGCCCGCGG      120
TACCGCGTTT TTGCGCGCA AGGCGCGGTT TGGCGCGGAC GCGGTGAGG GCGCGCGCAA      180
GCGCGCGGCT ACCGTCGCGA GCGTGGCGG TGGCGCGGCG AACGGCGGTC TCGCGCGCGG      240
GCGCGCGGAC GCGGTGTTT GCGGTGCGG TGGCGAGGGC GCGTGGGTG GCGAGGGGCG      300
TATGCGCGG GCGTGGCGG GCGGTGCGG TGGCGGCGG GCGCGCGGCG GCGCGCGGCG      360
TATGCGCGG GCGTGGCGG GCGGTGCGG TGGCGGCGG GCGCGCGGCG GCGCGCGGCG      420
TATGCGCGG GCGTGGCGG GCGGTGCGG TGGCGGCGG GCGCGCGGCG GCGCGCGGCG      480
TATGCGCGG GCGTGGCGG GCGGTGCGG TGGCGGCGG GCGCGCGGCG GCGCGCGGCG      494

```

1. INFORMATION FOR SEQ ID NO:172:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:172:

GGGCGGCTGG TGCCGCGGGC CAGCTCTTCA GCGCCGGAGG CGCGGCGGGT GCGCTTGGGG	60
TTGCGCGCAC CGCGGCGCCAG GGTGGGGGTG GCGGTGCCGG AGCGGCGGGC GCGGACGCCC	120
CGCCGAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCGGGC GCGCTCGGGC	180
GCCGAGCGGG CAACGCCATT GCGGCGGCA TCAACGGCTC	220

(2) INFORMATION FOR SEQ ID NO:173:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 388 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:173:

ATGCGCGCAA CGGGGCCCCC GCGGTGCTG GCGGGGCCGG CCACTACAAT TTCCAACGGC	60
GGGCAGGCTG GTCCCGCGCG CCAAGGCGGC CAAGGCGGCG TCGCGGGGGC AAGCACCACC	120
TGATCGGCCT AGCCGCAACC GGGAAAGCCG ATCCAACAGG CGACGATGCC GCCTTCCTTG	180
CGGCTTGGGA CCAGGCGCGC ATCAGCTACG CTGACCCAGG CCACGCCATA ACGGCGGCCA	240
AGGCGATCTG TGGGCTGTGT GGTAAACGGC TAACAGGTCT ACAGCTGGTC GCGGACCTGC	300
GGGACTACAA TCCCGGGCTG ACCATGGACA GCGCGGCCCA GTTGGCTGCC ATGCGATCAG	360
TCCGCTACTG TCCCGAAGAC TTGGAACA	388

(2) INFORMATION FOR SEQ ID NO:174:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

CAAAAGGCGG CAGCGGCGGG CCGCGGCTGA AAGGCTGA TCGGCTGCTA TCGCGGGAAG	60
AGGCTGCTTA AGCGGCGACC CGCGGCGGCG CGCGGAAGGG CGGCGCGGGC GCGACCAAGT	120
TAATCGAAGG CGCGGACGCG AAGCGCGGCA AGCGGCTGA TCGCGGGGTC CGCGGCAAGG	180
TGGAAGAGCG CGGAAAGGCG GAGGAGAGCA TCAAGAGCGG CGCGGCGGGC AAGCAAGGTC	240

(D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:177:

```

AGCAGCGCTA CCGGTGGCGC CGGGTTCGCC GGGGGCGCGG GCGGAGAAGG CGGAGCGGGC      60
GGCAACAGCG GTGTGGGCGG CACCAACGGC TCCGGCGGGC CCGGCGGTGC AGGCGGCAAG      120
GGCGGCACCG GAGGTGCCGG CGGGTCCGGC GCGGACAACC CCACCGGTGC TGGTTTCGCC      180
GGTGGCGCGC GCGGCACAGG TGGCGCGGCC GCGCGCGGGC GGGCCGGCGG GGCAGCCGGT      240
ACCGCGCGCA CCGGCGGCGT TGTGCGCGCC ACCGGTAGTG CAGGCATCGG CCGGCGCGGC      300
GGCGCGCGCG GTGACGGCGG CGATGGGCGC AGCGGTGTCC GCCTGGGCGT CTCGGCGTTT      360
GACGGCGGCG AAGCGCGGCG AGGCGGCGCG GCGGCGAGCG CCGGCGCGCG CGGCATCAAC      420
GGGCGCGCGC GGGCGCGCGG CAACCGCGCG GACGCGCGCG ACGGCGCAAC CGGTGCCGCA      480
GTTGTGCGCG ACAACGCGCG GTTCGCGCGT GACGGTGGCG CCGGTGGCGC CGCGCGCAAC      540
GGGGCGCAAC CCGGCGTGGG CGTGACAGGC AAAGCGCGCG AGGGCGCGCG CCGGGGCAAT      600
GGCGCGCAAC GGGCGCGCGG CGGTGTGGCG GCGGCGCGCG ACAACAATTT CAACGGCGCG      660
GAGGGTGGTG GGGCGCGCGA AGGCGCGCGA GCGGCGTTTG CCGGCGCGCA CACCACCTGA      720
TGGGGTAGCG CGGACCTGGG AAAGCGGATG CAACAGCGCG CGATGCGCGC TTCCTTGCGG      780
GTTGCGAGCA GCGGCGCATC AGGTAGCGTG AGCGGCGCGA GCGGATAAGG GCGGCGAAGG      840
CAATGTGTGG GGTGTGTGGT AAGGCGGTAA GAGGTGTAGA GTGCTGCGCG GACCTGCGCG      900
GATAGCATCG GGGGCTGAGC ATGGATAGCG GCGGCAATTT GGTGCGGATG GCATCAGGCG      960
GTACTGCGCG CGAAGCGCTG GAACA                                     985

```

X2) INFORMATION FOR SEQ ID NO:178:

X2) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1134 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

X3) SEQUENCE DESCRIPTION: SEQ ID NO:178:

```

GAGGAGAGAG ATGGGTAGCG GCGGCGCATG GCAGGTGCGG ATTGATGTGG GTTCCGCGAG      60

```

GGGGGACGG TGCACCTCTCA GGCAGCACCG GTGGTGCCGG

400

(2) INFORMATION FOR SEQ ID NO:175:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 538 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GGCAACGGCG GCAACGGCG CATCGCCGGC ATTCGGCCGC AACGGCGTTC CGGGACGGGC	60
AGCGGCAACG GCGGCAACG GCGGCAGCGG CGGCAACGGC GGCAACGGCG GCATGGGCGG	120
CAACAGCGGC ACCGGCAGCG GCGACGGCGG TGCCGCGCGG AACGGCGCGG CGGCGGGCAC	180
GGGCGGCACC GCGGCGGACG GCGGCGTACG CGGTACTGGC GGCACCGCGG GCAGCGGTGG	240
CACCGCGCGT GACGCGCGTA ACGGCGGCAA GCGAGCAGAT AACACCGCAA ACATGACTGC	300
GCAGGCGCGC GGTGACGGTG GCAACGGCGG GACCGGTGGC TTGCGCGCGG GGGCCCGGGC	360
CGCGCGCGGT GCGTTGACCG CTGGCGGCAA GCGCACCGCG GGGCAAGGCG GCGCCCGCGG	420
CGATGGCGCG AACGGGCGCA TCGGCGGCGA GGGCGGCTC ACTGACGACC CCGGCGGCAA	480
GGGGGCGACC GCGGCGAACC GCGGCACCGG GCGCACCGCG GCGCGCGGTA TCGGCAGC	538

(2) INFORMATION FOR SEQ ID NO:176:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 239 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

AGGAGTAA AGGAGTAA AGGAGTAA AGGAGTAA AGGAGTAA AGGAGTAA	60
TTGGGTTT TGGGTTT TGGGTTT TGGGTTT TGGGTTT TGGGTTT	120
GGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA	180
GGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA	240
GGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA AGGTTTAA	300

(2) INFORMATION FOR SEQ ID NO:177:

[illegible]

CGCGGCGGGC GAGCAGCTGA CCGCTGCGTT ATCGGCCATG TCCCGCGCGA TGAACGAAGG 1920
 AATGGCCTAA GCGCATTTGTT GCGGTGGTAG CGACTACGCA CCGAATGAGC GCGGCAATGC 1980
 GGTCAATCAG CGCGCCCGAC ACGGCGTGAG TACGCAATTGT CAATGTTTTG ACATGGATCG 2040
 GCGGGGTTCTG GAGGGCGCCA TAGTCCTGGT CGCCAATATT GCCGCAGCTA GCTGGTCTTA 2100
 GGTTCGGTTA CGCTGGTTAA TTATGACGTC CGTTACCA 2138

(2) INFORMATION FOR SEQ ID NO:179:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 460 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
 1 5 10 15
 Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
 20 25 30
 Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln
 35 40 45
 Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
 50 55 60
 Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala
 65 70 75 80
 Tyr Gly Glu Val Asp Glu Gln Ala Ala Thr Ala Leu Asp Asn Asp Gly
 85 90 95
 His Gly Thr Val Gln Ala His Ser Ala His Ala Val His Gly Asn Ser
 100 105 110 115
 Pro Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Ala Pro
 120 125
 Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
 130 135 140
 His Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn
 145 150 155 160

Gln Trp Ile Leu His Met Ala Lys Leu Ser Ala Ala Met Ala Lys Gln
 195 200 205
 Ala Gln Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Glu His Pro
 210 215 220
 Thr Tyr Glu Asp Ile Val Gly Leu Glu Arg Leu Tyr Ala Glu Asn Pro
 225 230 235 240
 Ser Ala Arg Asp Gln Ile Leu Pro Val Tyr Ala Glu Tyr Gln Gln Arg
 245 250 255
 Ser Glu Lys Val Leu Thr Glu Tyr Asn Asn Lys Ala Ala Leu Glu Pro
 260 265 270
 Val Asn Pro Pro Lys Pro Pro Pro Ala Ile Lys Ile Asp Pro Pro Pro
 275 280 285
 Pro Pro Gln Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser
 290 295 300
 Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met
 305 310 315 320
 Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala
 325 330 335
 Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp
 340 345 350
 Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val
 355 360 365
 Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
 370 375 380
 Pro Ala Gly Ala Gly Asp Ile Ala Ile Leu Gly Gln Gly Arg Ala Gly
 385 390 395 400
 Gly Gly Ala Ala Leu Gly Ile Ile Val Ile Met Pro Met Gly Ala
 405 410 415
 Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu
 420 425 430
 Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Trp Thr Glu Ala Val Ile
 435 440 445
 Gly Asn Arg Arg Arg Gln Asp Ser Ser Ser Ser Ser Ser

- (A) LENGTH: 277 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

Ala Gly Asn Val Thr Ser Ala Ser Gly Pro His Arg Phe Gly Ala Pro
 1 5 10 15
 Asp Arg Gly Ser Gln Arg Arg Arg Arg His Pro Ala Ala Ser Thr Ala
 20 25 30
 Thr Glu Arg Cys Arg Phe Asp Arg His Val Ala Arg Gln Arg Cys Gly
 35 40 45
 Phe Pro Pro Ser Arg Arg Gln Leu Arg Arg Arg Val Ser Arg Glu Ala
 50 55 60
 Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Trp His Pro
 65 70 75 80
 Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala
 85 90 95
 Arg Asp Gln Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp
 100 105 110
 Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
 115 120 125
 Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val
 130 135 140
 His Arg Glu Asn Leu Val Ala Gly Arg Arg Val Phe Arg Val Lys Pro
 145 150 155 160
 Phe Glu Pro Asp Tyr Val Phe Ile Ser Arg Met Phe Pro Pro Ser Pro
 165 170 175
 His Val Gln Leu Arg Arg Ile Leu Ser Leu Leu Gly His Arg Ser Ala
 180 185 190
 Gln Phe Gly His Val Gln Tyr Pro Leu Ser Leu Leu Ile Gln Arg Ser
 195 200 205
 Leu Ala Ser Gly Ser Arg Ile Ala Phe Pro Val Val Lys Pro Pro Glu
 210 215 220
 Pro Leu Asp Val Ala Leu Gln Arg Gln Val Gln Ser Val Pro Pro Ile
 225 230 235

Cys Arg Phe Phe Glu Ile His Glu Val Gly Phe Thr Gly Arg Gly His
 260 265 270

Pro Arg Arg Ile Gly
 275

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:181:

Ala Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro
 1 5 10 15
 Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
 20 25 30
 Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
 35 40 45
 Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
 50 55 60
 Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
 65 70 75 80
 Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg
 85 90 95
 Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser
 100 105 110
 Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val
 115 120 125
 Asn Lys Asn Ile Ala Arg Cys Ala Ser Ile Tyr Arg Arg Leu Asn Arg
 130 135 140
 Ala Arg Ser Ser Ser Ser Ala Ala Arg Cys Arg Thr Ile Gly Thr Phe
 145 150 155 160
 Arg Arg Ser Arg Thr Ser Ala Ser Met Arg Val Ser Thr Asn Ser Pro
 165 170 175
 Glu Val Thr Ala Glu Val Ala Ser Gly Val Thr Arg Ala Thr Ser Gly

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 196 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

Gln Glu Arg Pro Gln Met Cys Gln Arg Val Ser Glu Ile Glu Pro Arg
 1 5 10 15
 Thr Gln Phe Phe Asn Arg Cys Ala Leu Pro His Tyr Trp His Phe Pro
 20 25 30
 Ala Val Ala Val Phe Ser Lys His Ala Ser Leu Asp Glu Leu Ala Pro
 35 40 45
 Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val
 50 55 60
 Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala
 65 70 75 80
 Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln
 85 90 95
 Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His
 100 105 110
 Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val
 115 120 125
 Gln His His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val
 130 135 140
 Gln Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His
 145 150 155 160
 His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Thr Arg Gly
 165 170 175
 Arg His Ser His Ala Gly His His Arg Arg Arg Arg Ala His Ser Val
 180 185 190
 Gly Gly Ser Ala
 195

2. INFORMATION FOR SEQ ID NO:183:

SEQUENCE CHARACTERISTICS:

(A) LENGTH: 196 amino acids

[illegible]

Arg Thr Gly Phe Val Val Ala Leu Val Leu Glu Ala Val Gly Leu Asp
 290 295 300

Arg Asp Val Ile Val Ala Asp
 305 310

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2072 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:184:

```

ATGCTGCGCA TTCCGCACGA GCTCAGCAGC CCAAGGGGCG GTTGGGCGAA CTCATCGAGG    50
CATTCGCGCA CGGGCTCGCC GCGAAGGGTA AGCAAATCAA CACCACGCTG AACAGCCTGT    100
CGCAGCGCTT GAACGCTTGA AATGAGGCGC GCGGCGACTT GTTCGCGCTG GTACCCACCC    150
TGGCGGTATT CCGAACGCG CTACATCAGG ACCACCAACA GTTCGTCGCG TTGAACAAGA    200
ACGTTGCGGA GTTCACCGAC AGGTTGACCC ACTCGGATCG GGACCTGTCT AACGCGATCG    250
AGCAATTGCA CAGCTTCTCG GCGTTCGCGC GCGCTTCTT GCGCAAGAAG CGCGAGGTGC    300
TGACCGATGA CCGCAATAAT CTGCGGACCG TGACCGACAC GTTCTGTGAG CCGGATCGCT    350
TGGATCGGTT CGAGACCGTC CTGCGATCTT TCGCGACCGT GCGCGCGAAT ATTAAGCAGC    400
TTTACGATCG CAGACACGCT CGCTGCTGTG GCGTTGCGCG GTTCACGAAAT TTGCGCAAGC    450
TGATCGAGTT CATCTGCGAG TCGAATCAGC CGCTAGCGCG GCTCGGTTAT CAAGAGCTCG    500
CGGAAGCTCT TCGCGCATAT CTGCGCGCAG TCGTGGATCG GATCAAGTTC AACTAGCTTC    550
CTTCTGCTGT GAAGCTGCGC AGGAGGCTGT CGACCTGCTT CAAAGAGATT CGTCTGCTGT    600
ATGCGCTGTT CGAGCGCGCG CAATCTGAGA CGGACAGCAI CTGCGCGCTT ATCTGCTGT    650
ATGATAGCGT GTTCTGACAC TCGAAGAGG AGGCGCTTGA CTGCTGCGCA CGCGGATCG    700
AGCGCTGTTA CTGCGGACCG ATCAGCGCAG GTTCTGCTGA CGCGGAGTTC CTGCGCGAAG    750
CTATGCGTGG TCGCGATATG GCGCTGCTGT CTGCAAGGCT CCAAGACCGCG CGCGGACCGCG    800
CAAGAGCTGA CAGCGATATG GCGCTGCTGT CTGCAAGGCT CCAAGAGCTTC CGAGAGCTTC    850

```

TCTTGTTCCT GTGCGCGGGC GTGCGGACCT TCCTGTTCGG GGTGTGATCT AGCCCGCGCC	1200
GTGGAACGAT GCGCGATCGG CACGTGTTGA TACCGGCGAT CACCGGCGTG GCCTTGATCG	1260
CGGCATTGCT CGCACATTCT TGSTACCGCA CAGAACATCC GCTCATAGAC ATGCGGCTTG	1320
TCCAGAACCG AGCGGTGCGG CAGGCCAACA TGACGATGAC GGTGCTCTCC CTCGGGCTGT	1380
TGGGCTCCTT CTTGCTGCTC CCGAGCTACC TCCAGCAAGT GTTGACCAA TCACCGATGC	1440
AATCGGGGGT GCATATCATC CCACAGGGCC TCGGTGCCAT GCTGGCGATG CCGATCGCCG	1500
GAGCGATGAT GGACCGACCG GGACCGGCCA AGATCGTGCT GGTGGGGATC ATGCTGATCG	1560
CTGCGGGGTT GGGCACCTTC GCTTTTGGTG TCGCGCGGCA AGCGGACTAC TTACCCATTC	1620
TGCGGACCGG CTTGGCAATC ATGGGCGATG GCATGGGCTG CTCGATGATG CCACTGTCCG	1680
GGCGCGGAGT GCAGACCGTG GCGGACATC AGATCGCTCG GGTTCGAGG CTGATCGCGG	1740
TCAACGAGCA GGTGGGCGGT TGATAGGGA CCGCACTGAT GTGGGTGCTG CTCACCTACC	1800
AGTTCAATCA CAGCGAAATC ATCGCTACTG CAAAGAAAGT CGCACTGACC GCAGAGAGTG	1860
CGCGCGCGCG GGGGGCGCGG GTTGACCTTT CTTGCTTACC GCGCCAAAGC AACTTGGCGG	1920
CGCAACTGCT GCATGACCTT TCGCAGCGCT AGCGCGTGGT ATTGCTGATA GCGACCGCGC	1980
TAGTGTGCTG GAGGCTGATC CCGCGCGCAT TCCTGCCGAA ACAGCAGGCT AGTCATCGAA	2040
GAGCACTGCT GCTATCGGCA TGACCTCTCG TT	2072

2. INFORMATION FOR SEQ ID NO.135.

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1923 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO.135.

TTTATTTTAA TAACTGTTTC TTGAATGCT TTGAATGCA TCGCTGTCTG ATGTTGAGAA	1
TATGCTGCTA GATTGAGGAA AATGAGGCGG TCAAGATGCG TGACGAGGAG CTCGCGCGTG	123
TGCTAGCTGT GGTGCACTTC TTGGGTGACA TCGAGAAAGT TGAGGAAGAA AAGCGCGAGG	183
CGGCTGAGGG GTTGGCGCGG AGGATTCAGT GAGAGAAAGC TAATCGGCA CCGACGATC	243
TTTATTTTAA TAACTGTTTC TTGAATGCT TTGAATGCA TCGCTGTCTG ATGTTGAGAA	303

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1055 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

```

CTGGCGTCCC AGTGTACCCG GCGATATGAC GTCGGCATTG AATTTGCGGG CCCC GCCCGGA      60
CCCGTCGCCA CCCAATCTGG ACCACCCGGT CCGTCAATTG CCGAAGGTCG CCAAGTCCGT      120
GCCCCAATGTC GTGCTGGGTT TCTTGAACGA AGGCCTGCCG TATCGGGTGC CCTACCCCCA      180
AACAACGCCA GTCCAGGAAT CCGCTCCCGC GCGGCCGATT CCCAGCGGCA TCTGCTAGCC      240
AGGCTGCTTT GAGACCTAAC GTTGGCTAG GTGAAACCG GCGCCAGGGC CGCTGGACCG      300
GCTCATGGCA GCGAAATTAG AAAACCCCGG ATATTGTCCG CGGATTGTCA TAGGATGCTG      360
ATGCTTTGCT GGTTCGTGTT TAGCCATTGA GTGTGGATGT GTTGAGACCC TGGCCTGGAA      420
GGGACACAAC TGCTTTTGCG TCTTGGTCCG CTTTTCGCCG CCGACGCGGT GGTGGCGAAA      480
CGGCTGAGT CGGGAATGCT CGGCGGGTTC TCGCTTCCCG TCAGCTGGGG AGTGGCTGTG      540
CGACCCGATG ATTATGACCA GTGGGCGGCT GCGCCCGAAG ACCGCGCCGA TGTGATGTC      600
GAGCGCGCCG AAGGGGCGGA GCGAGAGGCG GCGGCCATCG ACCAGTGGGA TGAGTGGCAG      660
GCTCGAAGC AGTGGATGGC TGAGAAAGCT GAACCGCGCT TTCAGGTGCC ACCGAGTAGC      720
AGCAGCGTGA TTCCGATTC TCGGCTGCTT GGTAGAGA GCGGGCCGAG ACTGTCTTA      780
CTTGACCACT GATCGGCGGT CTGGTCTTT CGGCGCGCG GTATGACAGC AGTCAATCTG      840
GATGACAAGT TACAGGTATT AGGTCCAGGT TCAACAAGCA GACAGGCAAC ATGGCAACAC      900
GTTTATGAG CGATCGGCA TCGATCGCA GAGTTCGCA GTTTTTCAG ATGCAAGGCG      960
GAGCTGGA GAGCAGGCT GCGGATCTT GATTTTTC TCAAAAGATG GCGGCGGCT      1020
GTCGATTC GATGCGGCA GCGATCTCT TAAAT      1055

```

(i) INFORMATION FOR SEQ ID NO:187:

(A) SEQUENCE CHARACTERISTICS:

- A LENGTH: 312 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

TCCGGGCTGA CCACCGGGAT CCGCGAACCA TCCGAGATCA CCTCGCAATG ATCCACCTCG 120
 CGCAGCTGGT CACCCAGCCA CCGGGCGGTG TCGCAGACCG CCTGCATCAC CTTGGTATAG 180
 CCGTCGCGCC CCAGCCGCGAG GAAGTTGTAG TACTGGCCCA CCACCTGGTT ACCGGGACGG 240
 GAGAAGTTCA GGGTGAAGGT CGGCATGTCG CCGCCGAGGT AGTTGACCCG GAAAACCAGA 300
 TCCTCCGGCA GGTGCTCGGG CCCGCGCCAC ACGACAAACC CGACGCCGGG ATAGGTCAG 359

(2) INFORMATION FOR SEQ ID NO:188:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 350 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:188:

AACGGGCCCCG TGGGCACCCG TCGTSTAAGG CCGTCTGTTG GTCCCATGAA CTGCTGGAAC 60
 GATGCATCTT GGCAGATTCC CGCCAGAGCA AAACAGCCCG TAGTCCTAGT CCGAGTCGCC 120
 CCGAAAGTTC CTGAATAAC TCGTACCCG GAGCGCCAAA CCGGCTCTCC TTCGCTAAGC 180
 TGGCGGAACC ACTTGAGGTT CCGGGACTCC TTGACGTCCA GACCGATTCC TTCGAGTGGC 240
 TGATCGCTTC CCGCGCTGG CGCGAATCCG CCGCCGAGCG GGTGATGTC AACCCAGTCC 300
 TTGCGCTGGA AGAGGTGCTC TACGAGCTGT CTGGATGGA GACTTCTCC 359

2. INFORMATION FOR SEQ ID NO:189:

1. SEQUENCE CHARACTERISTICS
 A. LENGTH: 672 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS:
 D. TOPOLOGY: linear

3. SEQUENCE DESCRIPTION: SEQ ID NO:189:

1. Ile Pro Lys Gly Arg Thr Gly Ala Val Ile Glu Ala Phe Ala Asp
 1 5 10 15
 16. Leu Ala Gly Lys Gly Lys Glu Ile Asn Thr Thr Ser Asn Ser Leu
 20 25 30
 31. Ile Glu Lys Asp Ala Leu Asn Ile Glu Arg Gly Asn Phe Phe Ala
 35 40 45

Gln Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Glu Phe Thr Asp Arg
 65 70 75 80
 Leu Thr His Ser Asp Ala Asp Leu Ser Asn Ala Ile Gln Gln Phe Asp
 85 90 95
 Ser Leu Leu Ala Val Ala Arg Pro Phe Phe Ala Lys Asn Arg Glu Val
 100 105 110
 Leu Thr His Asp Val Asn Asn Leu Ala Thr Val Thr Thr Leu Leu
 115 120 125
 Gln Pro Asp Pro Leu Asp Gly Leu Glu Thr Val Leu His Ile Phe Pro
 130 135 140
 Thr Leu Ala Ala Asn Ile Asn Gln Leu Tyr His Pro Thr His Gly Gly
 145 150 155 160
 Val Val Ser Leu Ser Ala Phe Thr Asn Phe Ala Asn Pro Met Glu Phe
 165 170 175
 Ile Cys Ser Ser Ile Gln Ala Gly Ser Arg Leu Gly Tyr Gln Glu Ser
 180 185 190
 Ala Glu Leu Cys Ala Gln Tyr Leu Ala Pro Val Leu Asp Ala Ile Lys
 195 200 205
 Phe Asn Tyr Phe Pro Phe Gly Leu Asn Val Ala Ser Thr Ala Ser Thr
 210 215 220
 Leu Pro Lys Gln Ile Ala Tyr Ser Gln Pro Arg Leu Gln Pro Pro Asn
 225 230 235 240
 Gly Tyr Lys Asp Thr Thr Val Pro Gly Ile Trp Val Pro Asp Thr Pro
 245 250 255
 Leu Ser His Arg Asn Thr Gln Pro Gly Trp Val Val Ala Pro Gly Met
 260 265 270
 Ala Glu Val Gln Val His Pro Ile Thr Gln His Leu Leu Thr Pro Gln
 275 280 285
 Ser Leu Ala His Leu Met Gln Gly Ser Asn Ile Ala Pro Tyr Ser Ser
 290 295 300
 Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Gln Tyr Pro
 305 310 315 320
 Val Leu Pro Pro Ile Gly Leu Gln Ala Pro His Val Pro Ile Pro Pro
 325 330 335

Thr Leu Ile Pro Ala Ala Phe Leu Pro Lys Gln Gln Ala Ser His Arg
 560 665 670

Arg Ala Pro Leu Leu Ser Ala
 675

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser
 1 5 10 15
 Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile
 20 25 30
 Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala
 35 40 45
 Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu
 50 55 60
 Arg Ala Lys Ile Glu Ser Gln Asn Pro Asp Ala Ala Arg Ala Asp Arg
 65 70 75 80
 Lys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala
 85 90 95
 Arg Ser Ala Arg Leu Ala Cys Arg Arg Leu Pro Ala Ser Val Pro Thr
 100 105 110
 Thr Arg Arg Asp Pro Arg Ile Arg
 115 120

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:191:

```

                20                25                30
Ile Ala Glu Gly Arg Gln Val Arg Ala Gln Cys Gly Ala Gly Phe Leu
   35                40                45
Glu Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser
   50                55                60
Pro Gly Ile Arg Ser Arg Ala Ala Asp Ser Gln Arg His Leu Leu Ala
   65                70                75                80
Gly Asp Gly Ser Asp Val Thr Val Gly
                        85

```

(2) INFORMATION FOR SEQ ID NO:192:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 119 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

x1. SEQUENCE DESCRIPTION: SEQ ID NO:192:

```

Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ser Thr Ala Leu Ala
 1                5                10                15
Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp
 20                25                30
His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
 35                40                45
Ile Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
 50                55                60
Pro Gln Glu Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
 65                70                75                80
Leu Val Gln His Gln His Arg Val Val Ala Ala Gln Val Val Asp Pro
 81                85                90                95
Leu Asp Gln Ile Leu Arg Gln Val Leu Ile Pro Ala Arg His Asp Lys
100                105                110
Pro Asp Ala Gly Ile Gly Gln
115

```

(3) INFORMATION FOR SEQ ID NO:193:

1. SEQUENCE CHARACTERISTICS:

X1) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu
 1 5 10 15

Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Gln Ser Lys Thr Ala
 20 25 30

Ala Ser Pro Ser Pro Ser Arg Pro Gln Ser Ser Ser Asn Asn Ser Val
 35 40 45

Pro Gly Ala Pro Asn Arg Val Ser Phe Ala Lys Leu Arg Glu Pro Leu
 50 55 60

Glu Val Pro Gly Leu Leu Asp Val Gln Thr Asp Ser Phe Glu Trp Leu
 65 70 75 80

Ile Gly Ser Pro Arg Trp Arg Glu Ser Ala Ala Glu Arg Gly Asp Val
 85 90 95

Asn Pro Val Gly Gly Leu Glu Glu Val Leu Tyr Glu Leu Ser Pro Ile
 100 105 110

Glu Asp Phe Ser
 115

2) INFORMATION FOR SEQ ID NO:194:

X1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 911 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:194:

TGCTACGCAG CAATGCTTTT GGTGACAGAT GTGGATGCCC GCGTGGCTGC TGCCGATGCC 60

ATGAAAGCCC TCGAGCTTTT TCGGCTTTT TGGGAAAGAA TCGAACTGCT TAAAGAGCTT 120

TGCGGCGCGG CGATGCTGCG GTGCTGCGA GAGTGCAGCT TCGGCTGCTT TCGAGAGCA 180

TGGGTGTTTG GTTGGCTTTT CGAGCTGCGA TCGGCTGCTT GGTGACGCGG GCGGCTGTTT 240

TGATGCGGTC GCGGCTGCGT GCGGCTTAA GCGCTGCGCG TCGGACGCTG GTGGGCTGCG 300

AGCGCTGCTT GCGGCGCGCG CAGCGGCGCA AGCGGCTGCT GCGACCGCGG TCGGAGCGCG 360

TGAGCTGCGG TGAAGTAA GAGTGGCG GTTGTGCG TGTGTGAT TGTGTGAT 420

TGGTGCTGGC	GTCTGCGATG	GTGGTTTACG	GGCAGGGGCG	CTATGACTGT	CCCCAGCATG	500
GACCGGTGGA	CCCGGTGCGG	CGGCGGCGAG	CCGACCTGGA	CAATGGGGTG	TTGAGACACC	650
GTTGCCCGGG	GTGCGGCGAG	CCAGTCATCT	GGCAATTGGT	CGACGAAGAT	GGCCCGTTGC	720
GGCCGCGCAG	CCTGTACGCG	GCAGCAAGAC	CGCGCAGGAG	CACTACGCGC	TGGCGTGGTC	780
GGAAACGAAT	GGCGGTTCGG	TGGTGGCGTT	G			811

(2) INFORMATION FOR SEQ ID NO:195:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 966 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:195:

GTGCGCGCAT	GTGGCGGAGT	ATGACTTTTCG	GGAAACACCGG	CGTAGTAGTC	GAAGATATCG	50
GACTTTGTGG	TCCCGGTGGG	GGGATAGAGC	AGCTGTGGGC	GTTGGTCAGC	GTGACCGGTT	120
GCTCGGACGC	CGAACCAGTG	CTTTCAACGT	AGCCTGTGGG	TCACACAAGT	TCGGAGCGTA	180
AGGTCACGGT	CAAAATATCG	GTGGAATTTC	GGCGTGAAGT	TCCGCTCGCG	GACAATCAAG	240
TCATACTCAC	TTACATCCGA	GGCATTGGGA	GGGTTTGGAT	GGCTTCGGCG	GTGGTGAACG	300
TGCGGCTCAA	GTGTACAGC	GCTACCGCAG	ACCACGACAT	CAGGTTCCAC	CAGGTGCACG	360
CGAAGGACAA	CGGACCCATG	GGTACAAAGT	GCCTGTGCGA	GGGTGTGGCG	GAGGTGCTCG	420
ACTACGCGCA	TCTTGGCGCG	GGTACGAGT	CGCGCGACGG	CCAAATGGTG	CGGATCACCG	480
AGGACGACAT	CGCCAGCTTG	GCTGAAGAAG	CGAGCGCGGA	GATCGAGGTG	TTGGAGTTCC	540
TGCGGCGCGG	TCAGCTGGAG	CGGATGATGT	TGGAACGCGG	GTACTTTTTG	GAGCGTGATT	600
CGAAGTCTTG	GAATTTGATG	CTGTGCTGG	CGAAAGCAAT	CTGTGAGAGT	GAAGGATGCG	660
CGATCTTGGA	TTGTTGCAAT	GTGCTGCAAT	CGAAGAAAT	CGAAGGCTTG	GATGCAAGAT	720
TGCGGCTGCA	GCTGCGGTAG	GACAAAGCGT	AGAACGATCG	AGACATTGCG	GAGCTGAAGT	780
GCGCGCTGAT	AGAAAGCGCT	GTGCGCGATT	ATGAAACCGA	AAATACCGTT	ACTGATGCGA	840
TGCGGCTGCG	TCAGCTGATG	CGACCTTTTT	CGGAGCTTGC	AGCGGCTGCG	TGCGGAGCTG	900
GAATGCGGAT	CGAAGGAGAT	CTGTGCTGGA	CGAAGCTTGC	GGCGGCTGCG	GTGATGCTTG	960

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2367 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

```

CCGCACCGCC GGCAATACCG CCAGCGCCAC CGTTACCGCC GTTTGCGCCG TTGCCCCCGT      50
TCCCGCCCCG CCGCGCGGCC CCGCCGATGG AGTTCTCATC GCCAAAAGTA CTGGCSTTGC      100
CACCAGAGCC GCGGTTGCGG CGGTACCGCC CAGCCCCGCC GACTCCACCG GCCCCACCGA      150
CTCGCCCCGT GCGACCGTTG CCGCCGTTGC CGATCAACAT GCGGCTGGCG CCACCGTTGC      200
CACCACGCGC AGCGGCTCCG CCGACCGGCC CGACACCAAG CGAGCTGCGG CCGGAGCCAC      250
CATCACCAGC TACCGCCACCG ACCGCCCCAGA CACCACCGAC CCGGCTGTTG TGAACGCTCG      300
CGGTGCGAGC AGCGCGCGCG TTACCGCGAA CCGCAGCGCG AACGCGCGCG CCGCCATCGC      350
CGCGCGCGCG GCGGTTGCGG CGGTGCGCGC CGTTGCGCGA CAACAACCGG CCGCGCGCGC      400
CGGTGCGCGC CCGCGCGCGG GTGCGCGCGG CGCGCGCGAC GCGAAGGCGG CTGCGCGCGT      450
TGGCGCGATG AGCAGCGTTG CCGCGCGACG CATCGGTTG TGCGTGGGGG TGTGGGCTGT      500
GAAACGTCGC GATCGCGAGC TTGCGCGCGC TTGCGCGCGG CCGCGCGCGT CCGCGCGTAC      550
AAGCGATAGC AGCGCGCGCA CCGCGCGCAT GTTTCGCGCG ATCAGCGAAT AGCGAGCGCG      600
CGCGCGCGAC ATTGCGCGCA CTGCGCGCTG GTGACGTTG CGCGCGCGAG CCGCGCATG      650
AAGCGCGGTT ACCACCGAAA CCGCGCGTAC CACCGCGTGA GTGCGCGAGT CGGATGTGTA      700
CGAAGCGCGC CGGTGCGCGG CCGCGCGTAC CAGCGCGACT CGCGCGCGGT ACACCGTTG      750
CGCGTTGCGC AAGCGAGT GCGAAGCGG CGGCGATG GCGCGCGG AAGCGCGG      800
CGCGCGGTTG CCGCGCGCGG CCGCGCGCA CCGCGTGA GCGCGTGA CCGCGTGA      850
CGTTCGCGTT CCGCGCGCGT CCGCGCGCGG CGTTCGCGT CCGCGCGGTT CCGCGCGTGA      900
CGTTCGCGCG AGTTCGCGT CCGCGCGGTT CGTTCGCGG CCGCGTTTGA TCACCGATG      950
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGCGG CCGCGCGGTT CGATTGACCG      1000
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1050
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1100
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1150
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1200
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1250
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1300
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1350
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1400
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1450
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1500
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1550
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1600
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1650
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1700
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1750
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1800
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1850
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1900
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      1950
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2000
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2050
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2100
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2150
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2200
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2250
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2300
CGTTCGCGAT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT TCGCGCGGTT      2350

```

```

CGCCGGTACC ACCGGGCCCC CGCTTGCCCG CGTTGCCGAT CAACCCGGCG GCGCCTCCGC 1440
TGCCTGCGGT TTGACCGAAC GCGCCAGCCG CGCCGTTGCC ACCGTTGCCA AACAGCAACC 1500
CGCCGGGCGC GCGAGGCTGC CCGGGTGCCG TCCCGTCGGC GCGGTTTCCG ATCAACGGGC 1560
GCCCCAAAAG CGCCTCGGTG GCGGCATTCA CCGCAGCCAG CAGACTCCGC TCAACAGCGG 1620
CTTCAGTGCT GGCATACCGA CCGCGGGCCG CAGTCAACGC CTGCACAAAC TGCTCGTGAA 1680
ACGCTGCCAC CTGTACGCTG AGCGCCTGAT ACTGCCGAGC ATGGGCCCCG AACAAACCCG 1740
CAATCGCCGC CGACACTTCA TCGGCAGCCG CAGCCACCAC TTCCCTCCTC GGGATCGCCG 1800
CGGCGCGATT AGCGCGCTC ACCTGCGAAC CAATAGTCSA TAAATCCAAA CCGCGAGTTG 1860
CGACCACTC CGCCCTCCCG ATCACCAGG ACAGTTGCA CCTCCGATA CCCCATATCG 1920
CGCGAGCTC TCCCGAGCGG CCACGTGAGC TTGCTCGCT GGTGCGCGC CTGACTATC 1980
CGCGCGAGCG CCTCTCTCT GATTGCGCGG GCGCGCAGC TTGTTGCGC AGTTGAAGAC 2040
GGGAGGACAG CCGGAGCTTG GTGTAGACGT GGTCAAGTG GGAATGCAG GTCCGCGCGG 2100
AGATGAATAG CCGGAGCGCG ATCTCTCTCT TGTGAGTCC CTCACCGAGC AGTAGAGCCA 2160
CTCAAGCTC TGTGGTCTC AACCGCGCGG AGCCACTTCT CCGGCGTTT CTGCACTCC 2220
GGCTCTCTG CGGTAGCGG ATCGCTCAT CATCGATAA CGCAGTCTCT TCGGCGCAG 2280
CATCTCGAA CTGGCTCTCA CCGATCGATT TTGGAAGGT CGGTAGCGAC CAGTTACAG 2340
CGCTCTCTA CATCGCGAG CGGAGCG 2357

```

1. INFORMATION FOR SEQ ID NO:137

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 175 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION (SEQ ID NO:137)

```

Gln Pro Ala Gly Ala Thr Gly Ala Ala Ser Ser Pro Gly Ala Thr Val
1           5           10           15
Gly Ala Gly Gly Thr Gly Ser Pro Val Thr Thr Gly Thr Ala Ala
2           5           10           15
Thr Thr Gly Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
3           5           10           15

```

Val Ala Thr Ile Thr Ala Lys Gly Ala Arg Asn Val Ala Leu Arg Asp
65 70 75 80

Ser Ala Val Ala Ala Val Ala Ala Ala Thr Gly Ser Gly Gly Thr
85 90 95

Ala Val Thr Thr Gly Thr Ala Gly Gly Leu Ala Arg Ala Cys Arg Arg
100 105 110

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala
115 120 125

Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly
130 135 140

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly
145 150 155 160

Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly
165 170 175

Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr
180 185 190

Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
195 200 205

Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
210 215 220

Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
225 230 235 240

Ala Ser Ala Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala
245 250 255

Thr Val Ala Cys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys
260 265 270

His Thr Val Gly Ala Val Ser Asp Gln Arg Ala Thr Gln Cys Arg Leu
275 280 285

Ala Gly Arg Thr His Arg Thr Thr Gln Thr Pro Leu Asn Ser Gly Pro
290 295 300

Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu
305 310 315 320

Leu Val Lys Arg Thr His Leu Tyr Ala Thr Arg Leu Ile Leu Pro Ser
325 330 335

Arg Ser His His Phe Arg Arg Arg Asp Arg Arg Gly Arg Ile Ser Arg
 355 360 365

Ala His Leu Arg Thr Asn Ser Arg
 370 375

(2) INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2852 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:198:

```

33CCAAAACG CCGCGGGGAT CCGCGGCAAC CCGCGGGGCT ACCACCAAGT CTGGGCCCAG      60
GACGTGGCGG CGATGTTTGG CTACCATGCC CGCGCTTCGG CGGCGGTCTC CGCGTTGACA      120
CGGTTCGGCC AGGCGCTGCC GACCGTGGCG CGCGCGGCTC CGCTGCTCAG CGCGGCGCGC      180
GCTCAGGTGA CCACGCGGGT CTTCGCAAC CTGGGCTTGG CGAACCTCCG CGAGGGCAAC      240
GTCCGCAACG GTAATGTCG GAAGTTCAAT CTGGGCTCGG CCAACATCGG CAACGGCAAC      300
ATCGCGAGCG GCAACATCGG CAGGTCCAAC ATCGGCTTTC GCAACGTGGG TCCTGGGTTG      360
ACCGCGAGCG TGAACAACAT CGGTTCGGG AACACCGGCA GCAACAACAT CGGGTTTGCC      420
ACACCGGCGA GCAACAACAT CGGTTCGGG AATACCGGAG ACACCGAGCG AGGTATCGGG      480
CTACCGGCTA CGGTTCGTT CGGTTCGGG CGGTTCGAAT CGGCGAGCGA AACATCGGT      540
CTGTTCGAAT CGGCGAGCGG AACGTTCGG ATCGCGAAT CGGTATCGG GAATCGGGG      600
ATTGCGAAT CGGCGAGCG CTACACACG CGTTTCGCA ACTCGCGCA CGTCAACAG      660
CGGTTCGCA ACTCGCGAAT AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      720
CTACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      780
CTACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      840
CTACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      900
CTACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      960
AAACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      1020
CTACACACG ACTCGCGCA AGCGACAGT CGGTTCGCA ACTCGCGCA CTACACACG      1080

```

TTGGGTGGGG CACGGGTCTT CAATGTGGG CTGGCAAAAG GGGGGGTGGT GAACATTCTC	1250
GGCAACGGCA ACATCGGCAA TTACAACATT CTCGGGAGCG GAAACGTCCG TACTTCAAC	1320
ATCCTTGGCA GCGGCAACCT GGGCAGCCAA AACATCTTGG GCAGCGGCAA CGTCCGCAGC	1380
TTCAATATCG GCAGTGGAAA CATCGGAGTA TTCAATGTGG GTTCCGGAAG CTTGGGAAA	1440
TACAACATCG GATCCGGAAA CCTCGGGATC TACAACATCG GTTTTGGAAA CGTCGGCGAC	1500
TACAACGTGG GCTTCGGGAA GCGGGGCGAC TTCAACCAAG GCTTTGCCAA CACCGGCAAC	1560
AACAACATCG GTTTCGCCAA CACCGGCAAC AACAACATCG GCATCGGGCT GTCCGGCGAC	1620
AACGAGCAGG GCTTCAATAT TGCTAGCGCG TGGAACTCGG GCACCGGCAA CAGCGGCTC	1680
TTCAATTCCG GCACCAATAA CGTTGGCATC TTCAACGCGG GCACCGGAAA CTCCGGCATC	1740
TCAAACTCGG GCACCGGAAA CTGGGTATC GGAACCGCG GTACCGACAA TACCGGCATC	1800
TTCAATCTC GCAGCTACAA CACGGGCATC CTCACGCGG GCGACTTCAA CACGGGCTC	1860
TACAACACGG GCAGCTACAA CACGGGCATC TTCAACGTG GTACACCAA CACCGGCAAC	1920
TTCAACGTG GTACACCAA TACCGGCATC TACAACCGG GTACACCAA CACCGGCTC	1980
TTCAATCTC GCAACGTCAA TACCGGCTC TTCAACACGG GCGACTTCAA CAATGGCTC	2040
TTGGTGGGG GCGATAACCA GGGGAGATT GCGATGATC TCTCGGTAC CACTGCATC	2100
ATCGGATAA ATGAGCAGAT GGTCAATGAC GTACACGAG TACTGAGCTT GCGCGCAAC	2160
ATGATGAGG TACGAGAGG GTCAATGAT TTGGGCAAA GCTTGTATC TACGGGCTC	2220
TTCTCTTTC GCGGGGTCAA GGTCAATGAC TCAAGCTCA GCTTGTATC TACGGGCTC	2280
AGGATCGGG GAGGAGCGT TACGAGCTC ATGAGCTTC TCGGTGCTC TACGAGCGGC	2340
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2400
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2460
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2520
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2580
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2640
ATGATGAGT TACGAGAGT TACGAGCTC TCAAGCTCA TACGAGCTC TACGAGCTC	2700

GTAAGCGAAT AAACCGAATG GCGGCCTGTC AT

2852

(2) INFORMATION FOR SEQ ID NO:199:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 943 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

Gly Gln Asn Ala Pro Ala Ile Ala Ala Thr Glu Ala Ala Tyr Asp Gln
 1 5 10 15
 Met Trp Ala Gln Asp Val Ala Ala Met Phe Gly Tyr His Ala Gly Ala
 20 25 30
 Ser Ala Ala Val Ser Ala Leu Thr Pro Phe Gly Gln Ala Leu Pro Thr
 35 40 45
 Val Ala Gly Gly Gly Ala Leu Val Ser Ala Ala Ala Ala Gln Val Thr
 50 55 60
 Thr Arg Val Phe Arg Asn Leu Gly Leu Ala Asn Val Arg Glu Gly Asn
 65 70 75 80
 Val Arg Asn Gly Asn Val Arg Asn Phe Asn Leu Gly Ser Ala Asn Ile
 85 90 95
 Gly Asn Gly Asn Ile Gly Ser Gly Asn Ile Gly Ser Ser Asn Ile Gly
 100 105 110
 Phe Gly Asn Val Gly Pro Gly Leu Thr Ala Ala Leu Asn Asn Ile Gly
 115 120 125
 Phe Gly Asn Thr Gly Ser Asn Asn Ile Gly Phe Gly Asn Thr Gly Ser
 130 135 140
 Asn Asn Ile Gly Phe Ile Asn Thr Gly Asn Ile Asn Arg Ile Ile Gly
 145 150 155 160
 Ser Thr Gly Ser Gly Leu Leu Ile Phe Gly Ile Leu Asn Ser Gly Thr
 165 170 175
 Ile Asn Ile Gly Leu Phe Asn Ser Gly Thr Gly Asn Val Gly Ile Gly
 180 185 190
 Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn Ser Gly Asn Ser Tyr
 195 200 205

225 230 235 240
 Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly
 245 250 255
 Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu
 260 265 270
 Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe
 275 280 285
 Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly
 290 295 300
 Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe
 305 310 315 320
 Asn Ser Gly Ala Gly Ser Ala Ser Gly Phe Leu Asn Ser Gly Ala Asn
 325 330 335
 Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly
 340 345 350
 Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn
 355 360 365
 Thr Val Ser Gly Leu Phe Asn Met Ser Leu Val Ala Ile Thr Thr Pro
 370 375 380
 Ala Leu Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe
 385 390 395 400
 Pro Gly Gly Pro Pro Val Phe Asn Leu Gly Leu Ala Asn Arg Gly Val
 405 410 415
 Val Asn Ile Leu Gly Asn Ala Asn Ile Gly Asn Tyr Asn Ile Leu Gly
 420 425 430
 Ser Gly Asn Val Gly Asp Phe Asn Ile Leu Gly Ser Gly Asn Leu Gly
 435 440 445
 Ser Ala Asn Ile Leu Ile Ser Gly Val Val Ile Ser Asp Asn Ile Gly
 450 455 460
 Ser Gly Asn Ile Gly Val Phe Asn Val Gly Ser Gly Ser Leu Gly Asn
 465 470 475 480
 Tyr Asn Ile Gly Ser Gly Asn Leu Gly Ile Tyr Asn Ile Gly Phe Gly
 485 490 495
 Asn Ala Gly Asn Tyr Asn Val Ala Phe Gly Asp Ala Ile Asn Thr Asn

Gly Asn Asn Asn Ile Gly Ile Gly Leu Ser Gly Asp Asn Gln Gln Gly
 530 535 540
 Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu
 545 550 555 560
 Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly
 565 570 575
 Asn Val Gly Ile Ala Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn
 580 585 590
 Pro Gly Thr Asp Asn Thr Gly Ile Leu Asn Ala Gly Ser Tyr Asn Thr
 595 600 605
 Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly
 610 615 620
 Ser Tyr Asn Thr Gly Gly Phe Asn Val Gly Asn Thr Asn Thr Gly Asn
 625 630 635 640
 Phe Asn Val Gly Asp Thr Asn Thr Gly Ser Tyr Asn Pro Gly Asp Thr
 645 650 655
 Asn Thr Gly Phe Phe Asn Pro Gly Asn Val Asn Thr Gly Ala Phe Asp
 660 665 670
 Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Gly Asp Asn Gln Gly
 675 680 685
 His Ile Ala Ile Asp Leu Ser Val Thr Thr Pro Phe Ile Pro Ile Asn
 690 695 700
 Gln Gln Met Val Ile Asp Val His Asn Val Met Thr Phe Gly Gly Asn
 705 710 715 720
 Met Ile Thr Val Thr Gln Ala Ser Thr Val Phe Pro Gln Thr Phe Tyr
 725 730 735
 Leu Ser Gly Leu Phe Phe Phe Gly Pro Val Asn Leu Ser Ala Ser Thr
 740 745 750
 Leu Thr Val Pro Thr Ile Thr Leu Thr Ile Gly Gly Pro Thr Val Thr
 755 760 765
 Val Pro Ile Ser Ile Val Gly Ala Ser His Ser Arg Thr Ile Thr Phe
 770 775 780
 Leu Lys Ile Asp Ile Ala Pro Gly Ile Gly Asn Ser Thr Thr Asn Pro
 785 790 795

```

Asn Val Gly Gly Gly Ser Ser Gly Val Trp Asn Ser Gly Leu Ser Ser
      820                      825                      830

Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly
      835                      840                      845

Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr
      850                      855                      860

Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly
      865                      870                      875                      880

Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn
      885                      890                      895

Ala Gly Leu Ala Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys
      900                      905                      910

Arg Ile Arg His Gln Leu Asp Thr Val Ser Thr Ile Ile Ser Ala Phe
      915                      920                      925

Cys Gly Ser Ala Ser Asp Gln Ser Asn Pro Gly Ser Val Ser Glu
      930                      935                      940

```

2 INFORMATION FOR SEQ ID NO:200:

(A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(E) SEQUENCE DESCRIPTION: SEQ ID NO:200:

GGATGATAT GGGGATCAT TATGTCATC AGGTGATCGA CATTGATCGG ACC

53

3 INFORMATION FOR SEQ ID NO:201:

(A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(E) SEQUENCE DESCRIPTION: SEQ ID NO:201:

AGTGAATTGA AGGCTGGGTT GGGGGGCTT CATTGTAAG GA

42

4 INFORMATION FOR SEQ ID NO:202:

(A) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:202:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T

31

2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTCTGAATTC AGCGCTGGAA ATCGTCGCGA T

31

3) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:204:

GGATCCAGCG CTGAGATGAA GACCGATCCC GCT

33

4) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 38 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:205:

GGATATCTTC AGAATTCAGG TTAAAGTTC ATTTCGGA

5) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 30 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:206:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CTTCATGGAA TTCTCAGGCC GGTAAGGTCC GCTGCGG

37

(2) INFORMATION FOR SEQ ID NO:208:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7676 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

TGGCGAATGG GACGCGCCCT GTAGCGCGCG ATTAAGCGCG GCGGGTGTGG TGGTTACCGG	60
CACCGTGACC GGTACACTTG CCAGCGCCCT AGCGCGCCCT CCTTTCGCTT TCTTCGCTTC	120
TTTTCTCGCC ACCTTCGCGG GCTTTGCGCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG	180
TTTCCGATTT AGTGGTTTAC GGCACCTGCA CCCCCAAAAA CTTGATTAGG GTGATGGTTC	240
ACCTAGTGGG CCATCGCCCT GATAGACGCT TTTTCGCCCT TTGACGTTGG AGTCCACGTT	300
CTTATATAGT GGACTCTTGT TCCAAATGCG AACACACTGC AACGCTATCT CGGTCTATTC	360
TTTGTATTTA TAAGGGATTT TCGCGATTTG CGGTCTATTC TTAAAAAATG AGCTGATTTA	420
ACAAAAATTT AACCGCAATT TTAAGCAAAAT ATTAAGCTTT ACAATTTGAG TTGGGACTTT	480
TGCGGCAAAAT GTGCGCGGAA CGGTCTATTT TTTATTTTTC TAAATACATT CAATATGTA	540
CGCTCTATTC AATTATTTCT TAAAGCAAACT ATCGAGGAT CAAATGAAAG TCGAATTTAT	600
CAATATGAGT ATTATGATA TAAATTTTTC TAAAGCAAACT TTTCTGTAAT TAAGGAGAAA	660
ATTAATGAGT TAAATTTTCT TAAAGCAAACT TAAAGCAAACT TAAAGCAAACT TAAAGCAAACT	720
ATTAATGAGT AATAGCAAACT ATTAATTTTC TAAAGCAAACT AATAGGTTTA TAAAGTGAAG	780
AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT	840
AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT	900
AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT AATAGCAAACT	960

TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA 1140
TAAATTCGCT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC 1200
CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG 1260
TCGCACCTGA TTGCCCCGACA TTATCGCGAG CCCATTITATA CCCATATAAA TCAGCATCCA 1320
TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTTCCCG TTGAATATGG CTCATAACAC 1380
CCCTTGATT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA 1440
CGTGAGTTTT CTTCCACTG AGCCTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA 1500
GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACCACC GCTACCAGCG 1560
GTGGTTTGT TTCCGGATCA AGAGCTACCA ACTCTTTTTG CGAAGGTAAC TGGCTTCAGC 1620
AGAGCCGAGA TACCAAATAC TGTCTTTCTA GTGTAGCCCT AGTTAGGCCA CCCTTCAAG 1680
AATCTCTAG CACCCCTAC ATACCTCGCT CTCTAATCG TGTACCACT GCTCTCTCG 1740
AGTGGCGATA AGTCTGTCT TACCGGGTGG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800
CAGCGCTCGG GGTGAACGGG GGTCTCTGCG ACACAGCCCA GCTTGAGCG AACGACCTAC 1860
ACCGAACTGA GATACCTACA GCCTGAGCTA TGAGAAAGCG CCACCTCTCG CGAAGGGAGA 1920
AAGCGCGACA GGTATCCGCT AAGCGCGAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT 1980
CCAGCGCGAA ACCGCTGCTA TCTTTATAGT CTTCTCGGCT TTGCTCACCT CTGACTTGAG 2040
CTCGATTTT TGTGATGCTG CTGAGCGGCG CGGAGCCTAT CGAAAAACCG CAGCAACCG 2100
CCCTTTTAC GCTTCTGCG CTTTCTCTCG CTTTTCTCG ACATGTTCTT TCTCTCGTTA 2160
TCCCTCATT CTCTGATAA CCGTATTAGC GCCTTTGAGT GAGCTGATAC CGCTCGCGCG 2220
CTCGAAGCA TCGAGCGCAG CAGCTAGT AGCGAGGAA CGGAGAGCG CTGATGCG 2280
CTTTCTCTG CTAGGATCT CTGCTTAT CTGAGGAA CATATGCTG ACTCTGCTA 2340
CTCTCTCTG CTATGCGCA CTCTGAGCG ACTATAGCT CCGCTATCGG TACCTGCTT 2400
CTCTCTCTG CTCTGAGAG ACCCGCGAG ACCCGCTGAG CTCTCTGAG GGGCTTCTCT 2460
CTCTCTCTG CTCTGAGAG CAGAGCTCT GAGCTCTCTG CTGAGCTGCA TCTCTGAGAG 2520
CTCTCTCTG CTATGCTA AACTCTCTG CTCTCTCTG TAAAGCTCT CAGCTCTCTG 2580

GGTCACTGAT GGCTCCCTGT AAGGGGGATT TGTGTTGATG GGGSTAATGA TACCGATGAA	2750
ACGAGAGAGG ATGCTCACGA TACGGGTTAC TGATGATGAA CATGCCCGGT TACTGGAACG	2820
TTTTSAGGGT AAACAACCTG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG	2880
TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC	2940
TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCCGCTTT CCAGACTTTA	3000
CGAAACACGG AAACCGAAGA CCATTTCATG TGTGCTCAG GTCGCAGACG TTTTGCAGCA	3060
GCAGTCGCTT CACGTCGCT CGCGTATCGG TGATTTCATC TGCTAACCG TAAGGCAACC	3120
CCGCGAGCCT AGCGGGCTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCGC	3180
TATGCGCGG ATAATGCGCT GCTTCTCGCC CAAACCTTTG GTGGCGGGAC CAGTACGAA	3240
GGCTTCAGCG AGGGCGTCCA AGATTCCGAA TACCGCAAGC GACAGGCCCA TCATGCTCGC	3300
GCTCCAGCGA AAGCGGCTCT CCGCGAAAT GACCCAGAGC GCTGCCGGCA CCGTCTCTAC	3360
GATTTGCATG ATAAAGAAGA CAGTCATAAG TCGCGCGAGC ATAGTCATGC CCGCGGCCCA	3420
CGCGAAGGAG CTGACTGGGT TGAAGGCTCT CAAAGGCATC GGTCCGAGATC CCGGTGCCTA	3480
ATGAGTGAGC TAAGTTACAT TAATTGCTT GCGCTCACTC CCGGCTTTCC AGTCGGGAAA	3540
TTTGTCTGT CAGCTGCATG AATGAATCGG CCAAGCGCGG GCGAGAGGCG GTTTGCGTAT	3600
TGGGCGCGAG GGTGTTTTT TTTTTCAGCA CTGAGAGGCG CAACAGCTGA TTGCGCTTCA	3660
TTTGTGCGC CTGAGAGAGT TCGAGCAAGC GTTCAAGGCT GTTTTGGCGG AGCAGGCGAA	3720
TATGCTGCTT GATGCTGCTT AACGCGCGGA CATAACATGA GTGTGTTTCC GTATGCTGT	3780
ATCCCACTAC CGAGATATCC ACAGCAAGCG GCAGTCCGGA TTCGGTAATC GCGGCGATTG	3840
TACGAGCGCG CATCTGATCG TTGGCAAGCA GCATGCAATG GCGAAGGATG CCGTATTTCA	3900
TAATTCGAT GTTTTGTGA AAACCGGACA TTTATTTGA TTGCGCTTCA ATTTGCGCTA	3960
TAATTCGAT TTGATTCGCA CTGAGATATT TATTCAGCTT ATTCAGATTT AGAAGCGGCG	4020
AGACAGAACT TAATGGGCGG GCTAACAGCG GCATTTGCTG CTGAGCGAAT GCGACCGAGT	4080
GTTCGAGCGG CAGTGGCTTA CGGTCTTCAT GCGAGAAAAA AATAGTCTTT ATGGGTGTCT	4140
TTTCAAGAGC ATCAAGAAAT AACGCGCGGA CATTAGTGA GCGAGCTTCC ATAGCGAATGT	4200
TATTCGCTT ATTCAGCGCA TAATTAATTA CCAAGCGATC TATTCGCTT ATTCAGCGCA	4260

GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTTGTG 4440
CCACCCGGTT GGGAAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCGCGGTTT 4500
TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCAG 4560
CATACTCTCC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT 4620
CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTTGC CCATTTCGATG GTGTCCGGGA 4680
TCTCGACGCT CTCCCTTATG CGACTCCTGC ATTAGGAAGC AGCCCAGTAG TAGGTTGAGG 4740
CCGTTGAGCA CCGCCCGCCG AAGGAATGCT GCATGCAAGG AGATGGCCGC CAACAGTCCC 4800
CCGGCCACGG GGCCTGCCAC CATACCCACC CCGAAACAAG CGCTCATGAG CCGGAAGTGG 4860
TGAGCCCGAT CTTCCTCATC GGTGATGTCC GCGATATAGG CGCCAGCAAC CGCAGCTGTG 4920
GCGCCGCTGA TGCCCGCCAC GATGCCCTCC GCGTAGAGGA TCGAGATCTC GATCCCGCGA 4980
AATTAATAGC ACTCACTATA GCGGAATTGT GACCGGATAA CAATTCCTCT CTAGAAATAA 5040
TTTGTGTTAA CTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACCTCAT 5100
CGACATCATC GGGACCAGCC CCACATCTTC GGAACAGGCC CGGGCGGAGG CGGTCCAGCC 5160
CGCCCGGGAT AGGCTGATG ACATCCGCTT CGCTGCTGTC ATTGAGCAGG ACATGCGCTT 5220
TGACAGCCCC GCGAAGATCA CCTACCGCAT CAAGCTCGAA GTGTCTTTCA AGATGAGGCC 5280
CGCCCAACCC AGGGGCTCGA AAGCACCBAG TGGTTGCTCT TAAACGCGCC CGGGCCGCGG 5340
TACTGTGCCC ACTACCCGCC CTCTCTGCTT GTGACCTTC TCGGAGACCC CTAGCACTCT 5400
CTCTACCCCC CTGTTCAAGC TGTGGGTCTT GCGTTTTCAC GAGAGCTATC TGAAGCTCAC 5460
GATCACCTCT CAGGGCACCC GTTCTGGTGC TGGATGCCC CAGGCCGCCG CGGGGACGGT 5520
AAGATGTCG CCGTCGACC CTATCTCTTC TGAATCTTAT ATGCTGCTTC AAGAGGCTT 5580
ATGGAATATC TCGTAAGCA TCTCTGTTCA TCACTTCAAT TAAGAGTCT TCTGATCTAC 5640
TGAGGCTCTC AAGCTCAAGC TAAAACTCTT TCGGCTTATC TACGAGGCGA TCACTCAAAAC 5700
TTGGAAGCAG CCGCAGATCC TTGCGCTCAA TCGGCTGCTC AAGCTGCTCC GAGCTGCTCT 5760
AGTTGCTCTC CAGGCTGCTC ACGGCTGCTC TGACAGCTTC TTCTTCACCC ACTAGCTCTC 5820
TATGGAATAT TCGAGTCTT TCGGCTTATC TCACTTCTC GAGAGCTCTC TCACTTCTCT 5880

ACTCGGCGAG GCCCACTAG GCAATAGCTC TGGCAATTTG TTGTTGCCCC ACSCGCAAAAG	6060
CATTGAGGCC GCGCGCGCTG GCTTCGCATC GAAAACCCCC GCGAACCAGG CGATTTTCGAT	6120
GATCGACGGG CCCCCCCCCG ACGGCTACCC GATCATCAAC TACGAGTACG CCATCGTCAA	6180
CAACCGGCAA AAGGACGCGG CCACCGCGCA GACCTTGCA GCAATTTCTGC ACTGGGCGAT	6240
CACCGACGGC AACCAAGGCCT CGTTCTCTGA CCAGGTTTCAT TTCCAGCCCG TGCCGCCCCG	6300
GGTGGTGAAG TTGTCTGACG CGTTGATCGC GACGATTTTC AGCGCTGAGA TGAAGACCGA	6360
TGCGCTTACC CTGCGCGCAG AGGCAGGTAA TTTCGAGCGG ATCTCCGCGG ACCTGAAAAC	6420
CCAGATCGAC CAGGTGGAGT CGACCGCAGG TTCTTTGCAG GGCAGTGGC GCGGCGCGGC	6480
TGGGACGGC GCGCAGGCGG CGGTGCTGTC TTTCGAGAA GCGCGCAATA AGCAGAAGCA	6540
GGAATCTGAC GAGATCTGCA CGAATATTCC TCAGGCGCGG GTGCAATACT CGAGGGCGCA	6600
CGAGGAGCAG CAGCAGGCGC TGCTCTCGCA AATGGGCTTT GTGCGCACAA CGGCGCGCTC	6660
GCGGCGCTCC ACCGCTGCAG CGCCACCGCG ACCGCGGACA CCGTTTGCGC CCGCACACC	6720
GGCGCGCGCG AACACGCGCA ATGCGCGAGC GGGCGATCCC AACGCGAGCAC CTCCGCGCGC	6780
CGACCGCAAC GCACCGCGCG CAGCTGTGAT TGCGCGAAAC GCACCGCAAC CTGTCCGGAT	6840
CGACAACCGG GTTGGAGGAT TCAGCTTCCG CCGCGCTGCT GGCTGGGTGG AGTCTGAGCG	6900
CGCGCACTTC GACTACGCTT CAGTACTGCT CAGCAAAACG ACCGCGGAGC CGGCATTTCC	6960
TGGACAGCGG CCGCGCGTCC CGAATGAGAG CGGTATCGTC CTGCGCGCGC TAGACCAAAA	7020
CGTTTACGCG ACCCGCGAAG CGATCGATTC CAAGCGCGCG CCGCGCTTCC CCGCGAGAT	7080
CGGTGACTTC TATATCGGCT ACCCGCGCAC CGCGATCAAC CAGGAAACCG TCTCGCTTGA	7140
TGCGAAGCGG GTGTCTCGAA CGCGCTCGTA TTACGAGCTC AAGTTGAGCG ATCGAGTAA	7200
TGCGAAGCGG GAGATCTGCA CGCGCTGATC CGGTGCGCGC CGCGCGAAGC CAGCGGAGCG	7260
TGCGAAGCGG GAGCTGATCT GTGTCTGATC CTGCGGAGCT CGCGAAGAGC CGGTGAGCAA	7320
TGCGAAGCGG AAGCGGCTCG CGAATGAGT CGGTGCTTTG CTGCGCGCGC CGGTGCGCGC	7380
TGCGAAGCGT CTGCGAGAGC CGGTGCGCGC CGCGCGCGCG CGCGCGGAGG TCGTCTCTAC	7440
TGCGAAGCAA TGAGAGCTC AAGCGAGCTT AAGCGCTGCA GAATTTCTCA GATATCGATC	7500
TGCGAAGCGT CGGTGAGCTC AAGCGAGCTT AAGCGCTGCA GATATCTCTC GTAAAGAACT	7560

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

```

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser
 1             5             10             15

Pro Thr Ser Trp Glu Gln Ala Ala Ala Glu Ala Val Gln Arg Ala Arg
 20             25             30

Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
 35             40             45

Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
 50             55             60

Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
 65             70             75             80

Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
 85             90             95

Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
100            105            110

Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
115            120            125

Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
130            135            140

Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Thr Leu Ser
145            150            155            160

Glu Gly Asn Met Ala Ala His Lys Glu Leu Met Asn Ile Ala Leu Ala
165            170            175

Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His
180            185            190

Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile
195            200            205

```

Asp Thr Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly
 245 250 255
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val
 260 265 270
 Pro Gly Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys
 275 280 285
 Ala Glu Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp
 290 295 300
 Gln Ala Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser
 305 310 315 320
 Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Glu Ala Ala Ala Ala
 325 330 335
 Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
 340 345 350
 Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
 355 360 365
 Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
 370 375 380
 Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
 385 390 395 400
 Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
 405 410 415
 Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
 420 425 430
 Thr Leu Ala Gln Gln Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
 435 440 445
 Lys Thr Gln Lys Asp Gln Val Ile Ser Thr Ala Gly Ser Leu His Gln
 450 455 460
 Gln Thr Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
 465 470 475 480
 Phe Gln Gln Ala Ala Asn Lys Gln Lys Gln Gln Leu Asp Gln Ile Ser
 485 490 495
 Thr Asp Ile Arg His Ala Gly Thr Gln Tyr Ser Arg Ala Asp Gln Gln
 500 505 510

Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro
 530 535 540

Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro
 545 550 555 560

Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro
 565 570 575

Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Ile Asp Asn
 580 585 590

Pro Val Gly Gly Phe Ser Phe Ala Leu Pro Ala Gly Trp Val Glu Ser
 595 600 605

Asp Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr
 610 615 620

Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
 625 630 635 640

Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655

Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670

Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685

Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys
 690 695 700

Phe Ser Asp Pro Ser Lys Pro Asn Gly Ile Ile Trp Thr Gly Val Ile
 705 710 715 720

Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp
 725 730 735

Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asn Lys Ile Ala
 740 745 750

Val Tyr Ala Leu Ala Gln Ser Ala Asn Pro Leu Val Ala Pro Pro Pro
 755 760 765

Ala Pro Ala Pro Ala Pro Ala Glu Trp Ala Pro Ala Pro Ala Pro Ala
 770 775 780

Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu
 785 790 795 800

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 454 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

```

GTGGCGGCGC TGCGGCGGCG CAGCAGAGCG ATGTGCATCC GTTCGCGAAC CTGATCGCGG    60
TCGACGATGA GCGCGCCGAA CGCCGCGACG ACGAAGAACG TCAGGAAGCC GTCCAGCAGC    120
GCGGTCCGCG CGGTGACGAA GCTGACCCCG TCGCAGATCA GCAGCACCCC GCGGATGGCG    180
CCGACCAATG TCGACCGGCT GATCCGCCGC ACGATCCGCA CCACCAGCGC CACCAGGACC    240
ACACCCAGCA GGGCGCCGGT GAACCGCCAG CCGAATCCGT TGTGACCGAA GATGGCCTCC    300
TCGATCCGCA TCAGCTGCTT ACCGACCGGC GGGTGAACCA CCAGGCCGTA CCCGGGGTTG    360
TCTTCCACCC CATGGTTGTT CAGCACCTGC CAGGCCTGGC GGTGCGTAAT GCTTCTCCTC    420
GAAGATGGGG GTGCCGGCAT CCGTCACCGA GCCC

```

2. INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

```

TGCAGAAGTA TCGCGGATCC TCGGTGGCCG ACGCCGAACG GATTCGCCCG GTCCCGGAAC    60
GCATGCTCCG CACCAAGAAG GAAGGCAATG ACCTGCTCCT GTCTGCTCTT GCCATGGGGG    120
ATACCAACCGA CGACCTGCTG GATGTGGCTG AGCAGGTGTG CCCGGCGCCG CCGCCTCGGG    180
AGCTGGACAT GCTGCTTACC GCCGGTGAAC GCATGTCGAA TGCCTTGGTG GCCATGGGCA    240
TCGAGTCCGT TCGCCGCGAT GCCCGGTCTT TCACC3GTTT GCAGGCCCGG GTGATCACCA    300
TCCGCAACCA CCGCAACCGC AAGATCATCG ACCTCACGCG GGGGCGGCTG CAAACCGCCG    360
TTGAGGAAGG CCGGCTCCTC TTGGTGGCCG GATTCGAAGG GGTGAGCCAG GACACCAAGG    420
ATGTCACGAC GTTGGGCCCC GCGGGCTCGG ACACCACCGG CTTGCGCATG

```

3. INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 272 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CCCAGCTCCT CAAGGACGCG GAGAGCGATG AAGTCTTGGG CAAAATGAAG GTGTCTGCGC 240
TGGTTGAGGC GTTGCCAAAG GTGGGCAAGG TCCAGGCGC 279

(2) INFORMATION FOR SEQ ID NO:213:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ACACGGTCSA ACTCGACGAG GGCCTCGTGG AGGTGTGCGAC CGACAAGGTC GACACCGAAA 60
TCCCTCGCCG GCGCGGGGTG TGCTGAUCAA GATCATCGCC CAAGAAGATG ACACGGTCSA 120
AGTGGGCGCG GAGGTCTCTG TCATTGGGGA CGCCCATGAT GCGGGCGAGG CCGCGGTCCC 180
GGACCCCGAG AAAGTCTCTG CCGGCCCAAC TCGAATCCA 219

(2) INFORMATION FOR SEQ ID NO:214:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:214:

TGGCTGCGCA CATCGGCGCC GCGCCCGCCC GCAAGCGGCG ACCTAAGGCC GTCCCGGAGC 60
GAGCGCGGAC GCGGAAGGCC GAACCGGCAJ CATCGGCGCC GCGCGCGGAG CGAGCGGCTG 120
TGGCGGAGGG GGCACCGTAC GTGACGCGCC TGCTGCGAAA GCTGGCGCTG GAAAACAACA 180
TGGACCTGCG CCGGCTGACC GGCACCGGAG TGCTGCTGTC CATCGGCAAA GAGGATGTGC 240
TGGCGCGGCG TGAACAAAAG AAGCGGGCGA AAGCACCGCG GCGGGCGGCG GAGGCGGCGG 300
TGGCGCGGCG CCGGAAAGCG CCGCTGAAG ATCGGATGCG GC 342

(2) INFORMATION FOR SEQ ID NO:215:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:215:

CGGAAAACAA CATCGACCTC GCGGGGGTGA CCGGCACCGG AGTGGGTGGT CGCATCCGCA	300
AACAGGATGT GCTGGCCGCG GCTGAACAAA AGAAGCGGGC GAAAGCACCG GCGCCCTGAG	360
CGCTTCATCA CCGGTTAAC CAGCTTGCCC CAGAAGCCCG CTTCGACCTC TTCGCGGGTC	420
TTGGTCCGCT GCAGGCGGTC GCGGAGCCAG TTCAGGTTAG GCGGCCGAAA TCTTCCAGTT	480
CGCCAGGAAG GGCACCCGGA ACAGGGTCCG CACCC	515

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 557 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

XII SEQUENCE DESCRIPTION: SEQ ID NO:216:

TGGACCCCAA GGTGCAGATT CACACGCCA TTGAGGAAGC ACAGCGCACC CACCAAGCGC	60
TGACTCAACA GCGCGCCCAA GTGATCGGTA ACCAGCGTCA ATTGGAGATG CGACTCAAGC	120
GACAGCTGGC GGACATCGAA AAGCTTCAGC TCAATGTGCG CCAAGCCCTG ACGCTGGCCG	180
ACCAGGCCAC CGCGGCCGGA GACGCTGCCA AGGCCACCGA ATACAACAAC GCGCGCGAGG	240
CGTTGCGCAG CCAGCTGGTG ACCGCGGAGC AGAGCGTCGA AGACCTCAAG ACGCTGCATG	300
ACCAGCGCGT TAGCGCCGCA GCTCAGGCCA AGAAGCCCGT CGAACGAAAT GCGATGGTGC	360
TGCAGCAGAA GATCGCCGAG CGAACCAAGC TCTCAGCCA GCTCGAGCAG GCGAAGATGC	420
AGGAGCAGGT CAGCGCATCG TTGCGGTGCA TCACTAGGCT CGCCGCGCCA GGCAACACGC	480
CGAGCCTCGA CGAGGTGCGC GACAAGATCG AGCCTCGCTA CGCCAACGCG ATCGGTTCCG	540
CTGAACTTGC CGAGAGT	557

(2) INFORMATION FOR SEQ ID NO:217:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 223 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

XII SEQUENCE DESCRIPTION: SEQ ID NO:217:

AGGATAGCT TTGACATCG AGCTGGGTTT TCAAGGCT GCGGACCTT GTGATAGGCT	60
TAAGTGGAC TTCCGCGAG CAGGATGCA TTGAGGATG AACAGAAATG GCGTCTCGCG	120
TAAGATGCG GCGACTCGGT GAGAGGCTCA CCGAGGCGAG GATTAGGCG TTGCTCAAGC	180
AGGAAAGCGA CACCGTCCAA CTCGAGGAGT CCGTCTGGA GTT	223

2. INFORMATION FOR SEQ ID NO:218:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 179 base pairs
- B. TYPE: nucleic acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:218:

AAGAAGTACA	TCTGCCGGTC	GATGTCCGGC	AACCACGGCA	GCCAACCGGC	GCAGTAGCCG	60
ACCAGGACCA	CCGCATAACG	CCAGTCCCGG	CGCACAAACA	TACGCCACCC	CGCGTATGCC	120
AGGACTGGCA	CCCCCAGCCA	CCACATCGCG	GGCGTGCCGA	CCAGCATCTC	GGCCTTGACG	180
CACCACTGTG	CGCCGCAGCC	TGCAACGTCT	TGCTGGTCTG	TGGCGTACAG	CACCGGCCCGC	240
AACGACATGG	GCCAGGTCCA	CGGTTTGGAT	TCCCAAGGGT	GGTAGTTGCC	TGCGGAATTC	300
GTCCAGGCCCG	CGTGGAAAGT	GAACGCTTTG	GCGGTGTATT	GCCAGAGCGA	GCGCACGGCG	360
TCGGGCAGCG	GAACAACCGA	GTTGCGACCG	ACCGCTTGAC	CGACCGCATG	CCGATCGATC	420
GCGGTCTCGG	ACGCGAACCA	CGGAGCGTAG	GTGCCCAGAT	AGACCGCGAA	CGGGATCAAC	480
CCCAGCGCAT	ACCCGCTGGG	AAGCACGTCA	CGCCGCACTG	TTCCCAGCCA	CGGTCTTTGC	540
ACTTGGTATG	AACGTCCGCG	CGCCACGTCA	ACGCCAGC			578

(2) INFORMATION FOR SEQ ID NO:219:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 484 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:219:

ACAACGATCG	ATTGATATCG	ATGAGAGACG	GAGGAATCGT	GGCCCTTCCC	CAGTTGACCG	60
ACGAGCAGCG	CGCGGCCGCG	TTGGAGAAGG	CTGCTGCCCG	ACGTCCGAGC	CGAGCAGAGC	120
TCAAGGATCG	GCTCAAGCGT	GGCGGCACCA	ACCTCACCCA	GCTCCTCAAG	GACGCGGAGA	180
GGGATGAAGT	CTTGGGCAAA	ATGAAGGTGT	CTGCGCTGCT	TGAGGCTTTC	CCAAAGGTGG	240
GCAAGGTCAA	GGCGCAGGAG	ATCATGACCG	AGCTGGAAAT	TGCGCCCGAG	CCCGCCGCTT	300
TGGTGGCTTC	GGTGACCGTC	AGCCCAAGGC	TGTGCTGGAA	AAGTTCCGCT	CCGCTTAACC	360
CCCGCCGCTT	ACGATGCGGG	CCGGAAGGCG	TGTGCTGGGG	GTACCCCGCC	ATACGGGGGA	420
GAAGCCGCTT	GACAGGGCCA	GTCACAATT	TAGGCCGAA	CCCGGCTGG	GGGGGAACCC	480
CCCC						484

(2) INFORMATION FOR SEQ ID NO:220:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 537 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:220:

AAGAATGCA	CCGCAAGCCA	CCACATCCCG	GGCTGCCCGA	CCAGCATCTC	GGCCTTGACG	60
AAGAAATG	CCCGCAGCC	TCCAACTCT	TCTTGGTCTG	TGGCGTACAG	CACCGGCCCGC	120
AACGACATGG	CCAGGTCCA	CGGTTTGGAT	TCCCAAGGGT	GGTAGTTGCC	TGCGGAATTC	180

ACCTGCTACT GACGTCGCGC CGCCACGTCG AACGCCAGCG CCATCGCGCC GAAGAACAGC 480
ACGAASTACA CGCCGGACCA CTGGGTGGCG CAAGCCAATC CCAAGCAGCA CCCCAGC 537

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Gly Gly Ala Ala Ala Gly Gln Gln Ser Asp Val His Pro Phe Ala Asn
1 5 10 15
Leu Ile Ala Val Asp Asp Glu Arg Ala Glu Arg Arg Asp Asp Glu Glu
20 25 30
Arg Gln Glu Ala Val Gln Gln Arg Gly Pro Arg Gly Asp Glu Ala Asp
35 40 45
Pro Val Ala Asp Gln Gln His Pro Gly Asp Gly Ala Asp Gln Cys Arg
50 55 60
Pro Ala Asp Pro Pro His Asp Pro His His Gln Arg His Gln Asp His
65 70 75 80
Thr Gln Gln Gly Ala Gly Glu Pro Pro Ala Glu Ser Val Val Thr Glu
85 90 95
Asp Gly Leu Pro Asp Arg Asp Gln Leu Leu Thr Asp Arg Arg Val Asn
100 105 110
His Gln Ala Val Pro Gly Val Val Phe His Pro Met Val Val Gln His
115 120 125
Leu Pro Gly Leu Ala Val Arg
130 135

(2) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

His Lys Tyr Gly Gly Ser Ser Val Ala Asp Ala Glu Arg Ile Arg Arg
1 5 10 15
Val Ala Glu Arg Ile Val Ala Thr Lys Lys Gln Gly Asn Asp Val Val
20 25 30
Val Val Val Ser Ala Met Ile Asp Thr Thr Asp Asp Leu Leu Asp Leu
35 40 45

```

65          70          75          80
Glu Ser Leu Gly Ala His Ala Arg Ser Phe Thr Gly Ser Gln Ala Gly
          85          90          95
Val Ile Thr Thr Gly Thr His Gly Asn Ala Lys Ile Ile Asp Val Thr
          100          105          110
Pro Gly Arg Leu Gln Thr Ala Leu Glu Glu Gly Arg Val Val Leu Val
          115          120          125
Ala Gly Phe Gln Gly Val Ser Gln Asp Thr Lys Asp Val Thr Thr Leu
          130          135          140
Gly Arg Gly Gly Ser Asp Thr Thr Ala Val Ala Met
145          150          155

```

(12) INFORMATION FOR SEQ ID NO:223:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(13) SEQUENCE DESCRIPTION: SEQ ID NO:223:

```

Pro Ala Tyr Pro Ala Gly Thr Asn Asn Asp Arg Leu Ile Ser Met Arg
1      5      10      15
Asp Gly Gly Ile Val Ala Leu Pro Gln Leu Thr Asp Glu Gln Arg Ala
20     25     30
Ala Ala Leu Glu Lys Ala Ala Ala Arg Arg Ala Arg Ala Glu Leu
35     40     45
Lys Asp Arg Leu Lys Arg Gly Gly Thr Asn Leu Thr Gln Val Leu Lys
50     55     60
Asn Ala Glu Ser Asp Glu Val Leu Gly Lys Met Lys Val Ser Ala Leu
65     70     75     80
Leu Glu Ala Leu Pro Lys Val Gly Lys Val Gln Ala
85     90

```

(12) INFORMATION FOR SEQ ID NO:224:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(13) SEQUENCE DESCRIPTION: SEQ ID NO:224:

```

Thr Val Glu Leu Asp Glu Pro Leu Val Gln Val Ser Thr Asp Lys Val
1      5      10

```

[6]

Gly Asp Ala His Asp Ala Gly Glu Ala Ala Val Pro Ala Pro Gln Lys
 50 55 60
 Val Ser Ala Gly Pro Thr Arg Ile
 65 70

(2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 113 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

Ala Ala Asp Ile Gly Ala Ala Pro Ala Pro Lys Pro Ala Pro Lys Pro
 1 5 10 15
 Val Pro Glu Pro Ala Pro Thr Pro Lys Ala Glu Pro Ala Pro Ser Pro
 20 25 30
 Pro Ala Ala Gln Pro Ala Gly Ala Ala Glu Gly Ala Pro Tyr Val Thr
 35 40 45
 Pro Leu Val Arg Lys Leu Ala Ser Glu Asn Asn Ile Asp Leu Ala Gly
 50 55 60
 Val Thr Gly Thr Gly Val Gly Gly Arg Ile Arg Lys Gln Asp Val Leu
 65 70 75 80
 Ala Ala Ala Glu Gln Lys Lys Arg Ala Lys Ala Pro Ala Pro Ala Ala
 85 90 95
 Gln Ala Ala Ala Pro Ala Pro Lys Ala Pro Pro Glu Asp Pro Met
 100 105 110
 Pro

2. INFORMATION FOR SEQ ID NO:226:

i. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 118 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

ii. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:226:

Ala Leu Val Ser Ile Ser Ala Asp Glu Asp Ala Thr Val Pro Val Gly
 1 5 10 15
 Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Asp Ile Gly Ala Ala Pro
 20 25 30
 Val Pro Pro Ala Pro Lys Pro Val Pro Glu Pro Ala Pro Thr Pro
 35 40 45

```

65           70           75           80
Glu Asn Asn Ile Asp Leu Ala Gly Val Thr Gly Thr Gly Val Gly Gly
           85           90           95
Arg Ile Arg Lys Gln Asp Val Leu Ala Ala Ala Glu Gln Lys Lys Arg
           100           105           110
Ala Lys Ala Pro Ala Pro
           115

```

(2) INFORMATION FOR SEQ ID NO:227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:227:

```

Asp Pro Lys Val Gln Ile Gln Gln Ala Ile Glu Glu Ala Gln Arg Thr
 1           5           10           15
His Gln Ala Leu Thr Gln Gln Ala Ala Gln Val Ile Gly Asn Gln Arg
 20           25           30
Gln Leu Glu Met Arg Leu Asn Arg Gln Leu Ala Asp Ile Glu Lys Leu
 35           40           45
Gln Val Asn Val Arg Gln Ala Leu Thr Leu Ala Asp Gln Ala Thr Ala
 50           55           60
Ala Gly Asp Ala Ala Lys Ala Thr Glu Tyr Asn Asn Ala Ala Glu Ala
 65           70           75           80
Phe Ala Ala Gln Leu Val Thr Ala Glu Gln Ser Val Glu Asp Leu Lys
 85           90           95
Thr Leu His Asp Gln Ala Leu Ser Ala Ala Ala Gln Ala Lys Lys Ala
100           105           110
Val Gln Arg Asn Ala Met Val Leu Gln Gln Lys Ile Ala Glu Arg Thr
115           120           125
Lys Leu Leu Ser Gln Leu Glu Gln Ala Lys Met Gln Glu Gln Val Ser
130           135           140
Ala Ser Leu Arg Ser Met Ser Glu Leu Ala Ala Pro Gly Asn Thr Pro
145           150           155           160
Ser Leu Asp Glu Val Arg Asp Lys Ile Glu Arg Arg Tyr Ala Asn Ala
165           170           175
Glu Ala Gln Leu Ala Glu Ser
180           185

```

(2) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 71 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

Val Ser Thr Ser Thr Trp Val Pro His Pro Val Arg Asp Arg Val Ile
 1             5             10             15
Gly Gln Arg Trp Thr Cys Ala Asp Arg Arg Ser Ile Glu Glu Ser Thr
          20             25             30
Glu Met Ala Phe Ser Val Gln Met Pro Ala Leu Gly Glu Ser Val Thr
          35             40             45
Glu Gly Thr Val Thr Arg Trp Leu Lys Gln Glu Gly Asp Thr Val Glu
          50             55             60
Leu Asp Glu Pro Leu Val Glu
          65             70

```

(2) INFORMATION FOR SEQ ID NO:229:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:229:

```

Glu Val His Leu Pro Val Asp Val Gly Glu Pro Arg Gln Pro Thr Gly
 1             5             10             15
Ala Val Ala Asp Gln Asp His Arg Ile Thr Pro Val Pro Ala His Lys
          20             25             30
His Thr Pro Pro Arg Val Cys Gln Asp Trp His Arg Gln Pro Pro His
          35             40             45
Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala
          50             55             60
Ala Ala Cys Asn Val Leu Leu Val Asp Gly Val Gln His Arg Pro Gln
          65             70             75             80
Arg His Gly Pro Gly Pro Arg Phe Gly Phe Pro Arg Val Val Val Ala
          85             90             95             100
Cys Gly Ile Arg Gln Ala Arg Val Glu Val Gln Arg Phe Gly Gly Val
          105             110             115             120
Leu Pro His Arg Ala Val Thr Val Gly His Arg Asn Asn Arg Val Ala
          125             130             135             140
Thr Asp Arg Leu Thr Val Val Val Pro Thr Asn Arg Gly Leu His Arg
          145             150             155             160
Gln Pro Arg Ser Val Val Gly Gln His Asp Arg Gln Arg Asn Gln Pro
          165             170             175             180
His Arg Ile Pro Ala Gly Lys His Val Thr Pro His Cys Ser Gln Pro
          185
Arg Ser Leu His Leu Val
          190

```

(2) INFORMATION FOR SEQ ID NO:230:

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

```

Asn Asp Arg Leu Ile Ser Met Arg Asp Gly Gly Ile Val Ala Leu Pro
 1             5             10             15
Gln Leu Thr Asp Glu Gln Arg Ala Ala Leu Glu Lys Ala Ala Ala
 20             25             30
Ala Arg Arg Ala Arg Ala Glu Leu Lys Asp Arg Leu Lys Arg Gly Gly
 35             40             45
Thr Asn Leu Thr Gln Val Leu Lys Asp Ala Glu Ser Asp Glu Val Leu
 50             55             60
Gly Lys Met Lys Val Ser Ala Leu Leu Glu Ala Leu Pro Lys Val Gly
 65             70             75             80
Lys Val Lys Ala Gln Glu Ile Met Thr Glu Leu Glu Ile Ala Pro His
 85             90             95
Pro Ala Ala Phe Val Ala Ser Val Thr Val Ser Ala Arg Pro Cys Trp
100             105             110
Lys Ser Ser Ala Pro Pro Asn Pro Ala Gly Arg Arg Cys Gly Pro Glu
115             120             125
Gly Leu Trp Trp Ala Tyr Pro Arg Ile Arg Gly Arg Ser Gly Leu Thr
130             135             140
Gly Pro Ala His Asn Ser Gly Arg Thr Pro Arg Trp Gly Gly Thr Arg
145             150             155             160

```

(ii) INFORMATION FOR SEQ ID NO:231:

i. SEQUENCE CHARACTERISTICS:

A. LENGTH: 178 amino acids

B. TYPE: amino acid

C. STRANDEDNESS: single

D. TOPOLOGY: linear

ii. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:231:

```

Asn Lys His Arg Ile Pro Pro His Ala Leu Arg Ala Arg His His Leu
 1             5             10             15
Gly Leu Asp Ala Arg Leu Pro Ala Ala Ala Thr Asn Val Leu Leu Val
 20             25             30
Asp Gly Val Gln His Arg Pro Gln Arg His Gly Pro Gly Pro Arg Phe
 35             40             45
Gly Phe Pro Arg Val Val Val Ala Pro Gly Ile Arg Gln Ala Arg Val
 50             55             60
Ile Val His Arg Pro Gly Gly Val Val Pro Gln Arg Ala His Gly Val
 65             70             75             80

```

```

      115              120              125
Val Thr Pro His Cys Pro Gln Pro Arg Ser Leu His Leu Val Leu Thr
      130              135              140
Ser Arg Arg His Val Glu Arg Gln Arg His Arg Ala Glu Glu Gln His
      145              150              155              160
Glu Val His Ala Gly Pro Leu Gly Gly Ala Ser Gln Ser Gln Ala Ala
      165              170              175
Pro Arg

```

(2) INFORMATION FOR SEQ ID NO:232:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:232:

```

ATGCCAAGCC GGTGCTGATS CCCGAGCTCC SCGAATCGGT GACCGAGGGG ACCGTCATTC      60
GTTGGCTGAA GAAGATCGGG GATTCGGTTC AGGTTGACGA GCCACTCGTG GAGGTGTCCA      120
CCGACAAGGT GGACACCGAG ATCCCGTCCC CGGTGGGTGG GGTCTTGGTC AGTATCAGCC      180
CCGACGAGGA CCCCACGGTG CCCGTCGGCG SCGAGTTGGC CCGGATCGGT CTCGCTGCCG      240
AGATCGGGCG CCGCGCCGCG CCCAAGCCCT C

```

(2) INFORMATION FOR SEQ ID NO:233:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) MOLECULE TYPE: protein

(3) SEQUENCE DESCRIPTION: SEQ ID NO:233:

```

Glu Asp Phe Val Leu Met Pro Glu Ser Ile His Asp Val Thr Gln Trp
      10              20              30
Glu Val Ile Arg Trp Leu Leu Lys Ile Glu Asp Ser Val His Val Asp
      40              50              60
Glu Pro Leu Val Glu Val Ser Thr Asp Lys Val Asp Thr Glu Ile Pro
      70              80              90
Glu Pro Val Ala Gly Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala
      100              110              120
Thr Val Pro Val Gly His Glu Leu Ala Arg Ile Thr Val Ala Ala Glu
      130              140              150

```

(A) LENGTH: 107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

GAGGTAGCGG ATGGCCGGAG GAGCACCCCA GGACCGCGCC CGAACCGCGG GTGCCGGTCA 50
TCGATATGTG GGCACCGTTC GTTCCGTCCG CCGAGGTGAT TGACCAT 107

(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION. SEQ ID NO:235:

ATGAAGTTGA	AGTTTGCTCG	CCTGAGTACT	BCGATAGTGG	GTTGTGCAGC	GGCGCTTGTC	50
TTTCCTGCCT	CGGTTGCCAG	CGCAGATCCA	CGTAGCCGCG	ATCAGCCCGA	CATGACGAAA	120
GGCTATTGCC	CGGGTGGCCG	ATGGGGTTTT	GGCGACTTGG	CGGTGTGCGA	CGGCGAGAAG	180
TACCCGCAAG	GCTCGTTTTG	GCACCACTGG	ATGCAACCGT	GGTTTACCGG	CCCACAGTTT	240
TACTTCGATT	GTGTGAGCCG	CGGTGAGCCC	CTCCCGGGCC	CGCCGCCACC	GGTGGTTTGC	300
GGTGGGGCAA	TTCCGTCCGA	GCAGCCCAAC	GCTCCCTGA			33

A) LENGTH: 111 amino acids
B) TYPE: amino acid
C) STRANDEDNESS: single
D) TOPOLOGY: linear

XX. SEQUENCE IDENTIFICATION: SEQ. NO. 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844,

```
Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala
  1      5      10      15
Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp
  20      25      30
Pro His Glu Pro Asp Met Thr Lys Gly Thr Val Pro Leu Gly Arg Trp
  35      40      45
```

85 90 95
 Pro Gly Gly Cys Gly Gly Ala Ile Pro Ser Glu Gln Pro Asn Ala Pro
 100 105 110

2. INFORMATION FOR SEQ ID NO:237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

GTGACCAACGG	TGGGCGCTGCC	ACCAACCCGG	GCAGCGGCAG	CGGCGGCGGC	TCCGCGCGCT	50
CCGCGCGCAA	CGGTGGCGCC	GGGGGTAACG	CCACCGGCTC	AGGCGGCAAG	GGCGGCGCGC	100
GTGGCAATCG	CGGTGATCGG	AGCTTCGGCG	GTACCAAGCG	CCCCGCGCTC	ATCGGGGTCA	150
CGGCGCGCCC	CGGCGGCAAC	GGCGGCAAGG	GGCGCGCGCG	TGGCAGCAAC	CCCAACGGCT	200
CAGGTGGCGA	CGGCGGCAAA	GGCGGCAACG	CGGTGCGCGG	CGGCAACGGG	GGCTCGATCG	250
GGGCGAACAG	CGGCATCGTC	GGCGGTTCGG	GTGGGGCGCG	TGGCGCTGGC	GGCGCGCGCG	300
GAAACGGCAG	C					371

2. INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

CTCGGGGTCC	CACCAACCGG	CGGCGCGCGC	CTAGCGCGCC	GGCGCGACCA	CCCCCTTTTC	50
TTGACTCGTT	CAAGAAAAGG	GCCTTCTGTT	TGGTGGGCCA	TGTTGGCATT	ATCGTGACCC	100
ATGGGCAACA	TGACGTGCGA	ATCTCGCGCC	AAGGTCTAGC	TGCATGCGAA	TGGCGCGCGC	150
GGTGGTGAAT	ATGCTGTGAG	TCCTCATAGC	AGGGTTGCTG	GTACCTTTTC	CCGACCGACA	200
TCCTCGGAG	CCGCGCGCTG	TCTGCTACGC	GGTCTCTGGA	AAGGGGTTCG	TGGGCAACAT	250
ATCGCGCGCC	GCATATGGGT	GGGAGCGCGT	TTTACGAAAG	CTTTTCAAGG	TTTTTTGGGT	300
CGAATACCG	TCCTGCAACA	ATCGGCTGGA	ATCGCGCTG	CTGAGGCTGT	AAGACCATCC	350
CGAC						424

2. INFORMATION FOR SEQ ID NO:239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:239:

GCGATGGCGG	CCGCGGCTAC	CACCGCCAAT	GTGGAACGGT	TTCCCAACCC	CAACGATCCT	60
TTGCATCTGG	CGTCAATTGA	CTTCAGCCCC	GCCGATTTGG	TCACCGAGGG	CCACCGTCTA	120
AGGGCGGATG	CGATCCTACT	GCGCCGTACC	GACCGGCTGC	CTTTCGCCGA	GCGCGCGGAT	180
TGGGACTTGG	TGGAGTCGCA	GTTGCGCAGC	ACCGTCACCG	CCGACACGGT	GCGCATCGAC	240
GTCATCGCCG	ACGATATGCG	TCCCGAACTG	GCGGCGGCGT	CCAAACTCAC	CGAATCGCTG	300
CGGCTCTACG	ATTCGTC					312

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:240:

TCCCGTATGC	GCTTCGCAGC	CGGTGCGCGG	TCAACGCGCG	GGAGGCAATC	GCTTCGCTGC	60
CGACGAATGG	TTGCATCAGC	ATCGCAGTGT	GCGGTCGTGC	ACCGACACCG	CGGTCCAACG	120
TGAAGTCAGC	GCGGAAATC	GGCGGAAATC	TGCGCCCTCAG	TTGACGCTCG	CGCCCTAACG	180
GTTCTGGAAG	TTGGGTGCGC	GCTTCTCGGC	GAACGCGCGC	GGGCGTTCCG	TGGCGTCGTC	240
GGACAGGAAG	ACCTTGATGC	CGATCTGGGT	GTGCATCTTG	AACGCGCTCG	TTTCGGGCGT	300
CCACTCGGTC	TGCGGATGG	ACCGCAAGAT	GCGCTGCAGC	GCCAGGGGTC	CGTTAGCCGA	360
GATGGCGCTC	GCAAGTTCTA	GAAGTTGGT	CAACGCGCTG	CGGTGGGGCA	CACGTGGCCC	420
AT						422

(3) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 426 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:241:

TTTTTGGGTT	GAGGACGAGC	CGCGGGCTGT	TGATGTTGTT	GCACTCGGTA	GTCCGCGGCT	60
CGCGCTGCTT	CGTCTGCTCA	CGGGGCGCGG	GAGGATAAAG	GTCCCTAATG	CGCAGGTAGC	120
CGCGGAGGTC	CATGAGTCCG	ATGATGATGC	GACTGTGCGG	GTCCCGGAGC	CGCAGCTTGG	180
GATCGGCGCT	GATCAGGCGC	GACGCGTAGG	ATAAGTGGAT	CGAATGCATA	GTGGCGCTCA	240
GATGCGCGCT	GCGACTTCCG	CGGTGCTCCA	GCGGAAATG	CTTGATTTCT	AGCTCGCGCT	300
AGTGTTCGCG	CATGCGCTGT	GCGATGAATG	GGAATGCGA	GATCGGCGCA	AACGGGTCTG	360
AGCTGAGGTT	CGCGGCTTTT	CGCAGAGTGC	TGATATGCTG	GACTCGGCTA	GAAATGCTTG	420
ATCGCA						426

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

AGACCGGCGA GGGTGTGGTC GCTGCCCGCG GCATTGTGCGA TAATCTGCGC TGGGTGCGACG	60
CGCCGATCAA CTAGTGAGGC GCAACGCTAG GCTTTGGGAT ACCCACAGCT AAAAAGTTTA	120
TCAAAGAAAC GAAGAAGGTT GCCATGAGCA CTGTTGCCGC CTACGCCGCC ATGTGCGCGA	180
CCGAACCCCT GACCAAGACC ACGATCACCC GTCGCGACCC GGGCCCGCAC GACATGCGCA	240
TCGACATCAA ATTCGCCGGA ATCTGTGCGT CGGACATCCA TACCGTCCAA ACCGAATGGG	300
GGCAACCGAA TTTACCTGTG GTCCCTG	327

(i) INFORMATION FOR SEQ ID NO:243:

1. SEQUENCE CHARACTERISTICS:
A. LENGTH: 123 amino acids
B. TYPE: amino acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

ii. MOLECULE TYPE: protein

xii. SEQUENCE DESCRIPTION: SEQ ID NO:243:

Asp His Gly Gly Pro Ala Thr Asn Pro Gly Ser Gly Ser Arg Gly Gly	
1 5 10 15	
Ala Gly Gly Ser Gly Gly Asn Gly Gly Ala Gly Gly Asn Ala Thr Gly	
20 25 30	
Ser Gly Gly Lys Gly Gly Ala Gly Gly Asn Gly Gly Asp Gly Ser Phe	
35 40 45	
Gly Ala Thr Ser Gly Pro Ala Ser Ile Gly Val Thr Gly Ala Pro Gly	
50 55 60	
Ile Asn Gly Gly Lys Gly Gly Ala Gly Gly Ser Asn Pro Asn Gly Ser	
65 70 75 80	
Gly Gly Asp Gly Gly Lys Gly Gly Asn Gly Gly Ala Gly Gly Asn Gly	
85 90 95	
Ile Ser Ile Ile Ala Asn Ser Ile Ile Val Gly Ile Ser Gly Gly Ala	
100 105 110 115	
Ala Ala Gly Gly Ala Ile Gly Asn Ile Ser	
120 123	

(i) INFORMATION FOR SEQ ID NO:244:

1. SEQUENCE CHARACTERISTICS:
A. LENGTH: 104 amino acids
B. TYPE: amino acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

```

Met Ala Ala Ala Gly Thr Thr Ala Asn Val Glu Arg Phe Pro Asn Pro
 1           5           10           15
Asn Asp Pro Leu His Leu Ala Ser Ile Asp Phe Ser Pro Ala Asp Phe
      20           25           30
Val Thr Glu Gly His Arg Leu Arg Ala Asp Ala Ile Leu Leu Arg Arg
      35           40           45
Thr Asp Arg Leu Pro Phe Ala Glu Pro Pro Asp Trp Asp Leu Val Glu
      50           55           60
Ser Gln Leu Arg Thr Thr Val Thr Ala Asp Thr Val Arg Ile Asp Val
      65           70           75           80
Ile Ala Asp Asp Met Arg Pro Glu Leu Ala Ala Ala Ser Lys Leu Thr
      85           90           95
Glu Ser Leu Arg Leu Tyr Asp Ser
      100

```

(3) INFORMATION FOR SEQ ID NO:245:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. MOLECULE TYPE: protein

3. SEQUENCE DESCRIPTION: SEQ ID NO:245:

```

Ala Tyr Ala Leu Arg Ser Arg Cys Arg Val Asn Ala Pro Glu Ala Ile
 1           5           10           15
Ala Ser Leu Pro Arg Asn Gly Ser Ile Thr Ile Ala Val Cys Arg Arg
      20           25           30
Ala Pro Thr Pro Pro Ser Asn Val Asn
      35           40

```

4. INFORMATION FOR SEQ ID NO:246:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 25 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. MOLECULE TYPE: protein

3. SEQUENCE DESCRIPTION: SEQ ID NO:246:

```

Val Pro Leu Asn Thr Ser Pro Arg Leu Pro Arg Leu Pro Asp Ser Val
 1           5           10           15
Ala Ser Val Ala Ser Leu Leu Ser
      20

```

- (A) LENGTH: 61 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

```

Met Ser Thr Val Ala Ala Tyr Ala Ala Met Ser Ala Thr Glu Pro Leu
 1             5             10             15
Thr Lys Thr Thr Ile Thr Arg Arg Asp Pro Gly Pro His Asp Met Ala
 20             25             30
Ile Asp Ile Lys Phe Ala Gly Ile Cys Arg Ser Asp Ile His Thr Val
 35             40             45
Gln Thr Glu Trp Gly Gln Pro Asn Leu Pro Val Val Pro
 50             55             60

```

(ii) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

```

GCTTGGAGCC GTGGAGCCAC GGTGTGGGTG TGGGGGTGGA TTGCTTCTCG GCGAAAGTCA      60
ACTAAAGACC ACCTTGACAC GGAACGGGGC TCCCGCCATC GCGCGTCCGC GCTAGAGGC      120
TTTGACCGCC GCGCGAAGCC TTGCTGCTG TCGCGCATGC AGATCCGACA GCTTTGCTTC      180
AACATCGGGT GGAAGCGGTG TTAAGCCGAG GCT
                                         213

```

(ii) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

```

TTGAGCTGCT GTTGGGGGCT GCGGGTGGCT TCGGGGCGGG TCGGGGCGGC ATCGACGGCG      60
GCTTGGGTGC TACGGGCGGG AGCGGGGAG AAGGGGGAAT GCGGAGCGAC GTTGGATGGT      120
TTGCAATCGG TGGGGGCGGT GCGGGGCGGT TGAAGGCTG TGAAGTGTAT GTTGGAGCGG      180
ATGCTTGGCT GTTGGGTGCT AGCGGGGGA TGGGGGCTAT TCGGGGCGGC GTTGGGCGGT      240

```

(A) LENGTH: 420 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:250:

AAAGCGGTGAT	TGGCAAGGCG	ACCGCGCAGC	GGCCCGTAGC	CGCGGGACGG	CCCAGGCCCC	60
GACCGCAGCG	GCCGGTGTCT	GACCGGTCA	GGGACGAGG	GGCTGACCG	TGCCGCTCGT	120
CTACTTCGAC	GCCAGCGCCT	TGTCAAACT	TCTCACCAC	GAGACAGGA	GCTCGCTGGC	180
GTCCGCTCTA	TGGGACGGCT	GCGACGCCG	ATTGTCCAAC	CGCCTGGCCT	ACCCCGAAGT	240
CGCGCGCGCA	CTCGCTGCAA	CGGGCCGCAA	TCAGACCTA	ACCGAATGGG	AGCTCGCGCA	300
CGCGGAGCGT	GACTGGGAGG	ACTTCTGGG	GCGACCGCGC	GAGTGGAACT	GACCGCGACG	360
TTGAACAGC	ACGCGCGGCA	GCTCGCGGCA	ACACATGCCT	TACGCGGAGC	CGACACCGTT	420

(A) LENGTH: 299 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:251

TTGTTGTGCGG	TGGCATTCGGC	GGTACCGCGCG	GAACCGCGCGG	CAACCGCGCGT	ATGCTGCGCTG	179
CGCGCGCGCGG	GGCGCGCGCGT	GGCGCGCGCGT	TACAGTTTCAG	CAGTGGCGCGT	GGGGCTGCGCG	180
CGCGCGCGCGG	GGCGCGGTGGG	TTGTTTCACCA	CGCGCGCGTGT	CGCGCGCGCGG	GGTGGCGAGG	181
GTACACCGCGG	CGGGGCGCGCG	GGCGCGCGCGG	CGCGCGCGCGG	TTTGTTTTGGT	CGCGCGCGCGA	182
TGGCGCGGGG	CGCGCGGATTG	TGGGATGACCG	GAACCGCTCGG	CACCGCGCGG	CGCGCGCGG	183

```

A LENGTH 1000000000
B TYPE 1000000000
C STRANDEDNESS single
D TOPOLOGY linear

```

NO.	SEQUENCE DESCRIPTION	REV.	NO.	REV.
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87	87	87	87

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

Glu Leu Leu Phe Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly
 1 5 10 15
 Thr Asp Gly Gly Pro Gly Ala Thr Gly Gly Thr Gly Gly His Gly Gly
 20 25 30
 Val Gly Gly Asp Gly Glu Trp Leu Ala Pro Gly Gly Ala Gly Gly Ala
 35 40 45
 Gly Gly Gln Gly Gly Ala Gly Gly Ala Arg Ser Asp Gly Gly Ala Leu
 50 55 60
 Gly Gly Thr Gly Gly Thr Gly Gly Thr Gly Ala Gly Gly Ala Gly
 65 70 75 80
 Gly Arg Gly Thr Leu Leu Leu Gly Ala Gly Gly Gln Gly Gly Leu Gly
 85 90 95
 Gly Ala Gly Gly Gln Gly Gly Thr Gly Gly Gly Arg Arg Arg Trp Arg
 100 105 110
 Ser Gly Gly Cys Gln Trp His Trp Trp
 115 120

(2) INFORMATION FOR SEQ ID NO:254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Val Thr Gly Lys Ala Thr Ala Glu Arg Pro Val Ala Ala His Arg
 1 10
 Val Ala Glu Thr Gln Arg Pro Val Ser Asp Arg Val Ser Asp Gln
 21 30
 Arg Arg

(2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:255:

```

Leu Val Gly Gly Ile Gly Gly Thr Gly Gly Thr Gly Gly Asn Ala Gly
 1           5           10           15
Met Leu Ala Gly Ala Ala Gly Ala Gly Gly Ala Gly Gly Phe Ser Phe
      20           25           30
Ser Thr Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Leu Phe
      35           40           45
Thr Thr Gly Gly Val Gly Gly Ala Gly Gly Gln Gly His Thr Gly Gly
      50           55           60
Ala Gly Gly Ala Gly Gly Ala Gly Gly Leu Phe Gly Ala Gly Gly Met
      65           70           75           80
Gly Gly Ala Gly Gly Phe Gly Asp His Gly Thr Leu Gly Thr Gly Gly
      85           90           95
Ala Gly Gly

```

(2) INFORMATION FOR SEQ ID NO:256:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:256:

```

TTCCTGTTGGG CGCGCGCGCGG GTGGGCGGTH TTGGGCGGTGA CGGTGTGCGCA TTCCTGCGCA      60
TTCGCGCGCGG CGGCGCGCGGT GGTGCGCGGGG GGGCGCGGTGG GCTGTTGAGC CTCGGTGGGG      120
CGCGCGCGCGG CGCGCGCAATC GATTGCTGG GGAACAGCGG TCGCGCGCGG TCGCGCGCGGT      180
CGCGCGTGGT GTGGGCGGAG GCGGTGCGG GCGCGCGCGG TGGGCTGCGG TCGACTACG      240
CGCGTGGCGG CGCGCGCGCGG GGAACCGCA GCTGCTGCT AA      282

```

(2) INFORMATION FOR SEQ ID NO:257:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 415 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:257:

```

CGGACAGAGG CCGCTAATC GGAAGTGAAT GCGTGAATG GATTTTAT GCGCTGAAT      60
TATGCTGCTT GAGGAGGGA GGAATATTC AAAACCAATC TCGGCTGGA ATGAATATC      120
ATGCGCTGTA GCGCTGTAAT GCGCTGTAAT GCGCTGTAAT GCGCTGTAAT GCGCTGTAAT      180

```

ATGGGTGCCAG CCCACTCGAC ACCACCGGTG GCGAACATCG AGGTCAACAC GCCGT

415

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

TCACCGCGTG AACGGTTCGT AACACTGATA CGTATGCTTG TCAGCGAGCA GATCAAGTCC	50
AGTCCGACCA ATGCCAGGAG ATCATCGGCT AGGCTCACGG TTTCGCCTGG GACGAGACGG	100
CATTGAGTTC TGGCGTTGCA CGGTCCCTCC CTTGGTGGGA AGTCTGACGC GGCATCAGAA	150
CGTTGTCAA TACCACTCTT TGGGGGATAT GGCCTATTTC GTGTCTCCG GCGCGTCCAT	200
CGGATCCCTT TTCGAACCTT GCGCAAGCCC CGTCCAGTTA CGGCTGTTC ACTGCCGCT	250
CGCTAGCTG CGCGGCTCG ATCGGTTTGA ACCTCATCCC AATTCGCCCA ATGGGTGAGT	300
ACCTGACGCT CCT	373

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 403 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

CGAAACCCGA CAGGCCGGCA CGGACGGTCC GAAGTTGCAU AACGTTGGG GCTCCATCTA	50
CGGAACCGGT CACCACGGCG TAGACAGCAG ATCTGTGGAT CGGCGTTTC GTGTCTCCG	100
CGCCGAGTAC CGCGGGGCGG AACCGCAGCG ACCAAAGCAA CGCATCGAT ACGGGGATCG	150
CGACTCGTGC CGAATTGAG GTCCCTCGAC AAGCTTGGCG CGGACTCGA ACCCGGGTCA	200
ATGATTGACT TTAACCGCT TAGCAATAA TAGATAAGC GGTGGGGGG TGTAAACGGG	250
TTTGAAGGCC TTTTCTCTG AAAGGAGTAA TATATCGGG ATAACTGGT TAGTAGCGA	300
GAAGGCGCGG ACGCATCGG TTTCGAAAG TTTCGAAAG TTAATAGGG ATCGAGCGG	350
TTT	403

(2) INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

AGTGGCCAGC CGGTGGGCCA ATGCATCCAG GTCCCGGTAC GTCAGCTGAC CATCGGCCCA 60
ACTGACCGCC ACCGAGTCAG GCTGTGCCGC AGCGATTTCG GCGAACCGGG TATGCACCCC 120
GGGTGCCGAC GTCGTCACAT CCGGCAGGCC GGGTGCCGTC GGATCGTGCT CGCCGTCCAG 180
CAGAATGTCT ACGTCGCGCA GCGGCCGATC CCACCGGCTG ACCAAGCGCT GTAACACAGC 240
CAGCACCCGC CTGCCGAGGC TTTCGGGGCG CATCGTGCCC AGCGCACCGT CGAGCACCTC 300
CACTAGCAGC GTGACCTCAC CGGTGCTGCG GTGCCGCGCG ACGGTCACCG GAAAGTGCGA 360
CAAACCTCTT AGCGCCACCG GACGGAACCT CACCCCGTTT GCGA 404

```

(2) INFORMATION FOR SEQ ID NO:261:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 421 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: cDNA

12. SEQUENCE DESCRIPTION: SEQ ID NO:261:

```

GTCTGTCTCG CAGGTGTGTC TTGGAACCCG CTGGCTAACT TGGCACCCCG GTATCCGCCC 60
ACCATGGAAC CCGCCGAAGC GCGGGTGTCA CCGCTTACTT CGCAAGACCC GCGCGGTGCA 120
GTCCGACCAC TGAGCGGCGA CCGCCGGCCG GCACTATTTC ACAACGGCAC CCGCCAATTG 180
GTGCTCTGCG GCGCGGCGCG CGATTCCGCG GCACCGCGCA GCATCATGGT CTTCGATGAC 240
ATGCACTTTC CACCGCTCGT CATTTTCTG CCGGGCCCGG CAGCCCGGTT GACCAGCGAC 300
GACCAACCGA CGGCTTTCCT TCGCGCCCGC GCGCGCTACT TCGTGCGCGA CCGTCTCTTC 360
GCTCACACCG CACGAGTGAA TGTGCTGAC GCACCGCACA CGGATTTCAC CCGGATCGCC 420

```

2. INFORMATION FOR SEQ ID NO:262:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 426 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: cDNA

12. SEQUENCE DESCRIPTION: SEQ ID NO:262:

```

ATGTAATTA ATGTAACCG ATGCTGCTG GATATCTTC TGGCGGCGCG ATGTAATTA 60
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 120
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 180
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 240
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 300
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 360
ATGTAATTA ATGCTGCTG TGGCGGCGCG ATGTAATTA ATGCTGCTG TGGCGGCGCG 420

```

3. INFORMATION FOR SEQ ID NO:263:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:263:

GTCTTGGTGG	CAGGCTGTTC	TTCGAACCCG	CTGGCTAACT	TGCAACCCGG	GTATCCGCCC	60
ACCATCGAAC	CCGCCCAACC	GGCGGTGTCA	CCGCCTACTT	CGCAAGACCC	GGCCGGTGCA	120
GTGCGACCAC	TGAGCGGCCA	CCCCCGGGCG	GCATTAATCG	ACAACGGCAC	CCGCCAATTG	180
GTGGCTCTGC	GCCCGGGGCG	CGATTGGGCG	GCACCCGCCA	GCATCATGGT	CTTCGATGAC	240
GTGCACGTTG	CACCGCCCGT	CATTTTCTCG	CCGGGCCCCG	CAGCCGCGTT	GACCAGCGAC	300
GACCACGGCA	CGGCCCTTCT	TGCCGCCCCG	GGCGGCTACT	TGGTGGCCGA	CCTGTCTCTC	360
GGTCACACCG	CACGAGTGAA	TGTGCTGAC	GCAGCGCACA	CCGATTTTCA	CGCGATCGCC	420
CGCCGCTCCG	ACGGCAAGCT	GCTGCTGGGT	AGCGCAGATG	GCGCCGTCTA	CACGCTTGGC	480
AAGAACCCTG	AGTTGACCCG	GCTCGGCGCC	GCCACCGTAG	CC		521

2. INFORMATION FOR SEQ ID NO:264:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 739 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:264:

CTTGGGGGCG	ACCGCCCTGG	GGCGGCGGCA	CGCCCTGGGG	GCAGACCCCG	CGCAAAACCA	60
ACCGCTGGGG	CTTASTGGGG	GGCGGCGGCG	TGGTGGTGCT	GGTCTCTCTG	TTGGGGCGGCA	120
TGGGCATCTG	GATGGGATCT	TGGGCGGAGG	CGGTACAGCG	GGGTACGCGG	CTTGGCGGAGG	180
AGCGGCTTAG	CGCCCTAGCT	GTGAAGTCTT	TAGAAGTCAA	TGGCGTGATG	GGTCTCTCTG	240
AGTGGAGGCG	GGGCAAGCGG	ATGAGATGGA	TGGACTCTTG	GCGCGTGAGG	GTGCTCTCTG	300
TGGACTGCGA	GGGCGGCGCT	TATACGAGCG	AGGATCGCGT	GTATCGCGCG	ACCGGCTACA	360
GGGCGATCAA	CGGCTTGATG	TGATCGGAGG	CGGCGGACAA	GTACGAACAT	TGGGTGAAGC	420
AAGCGCTCTG	CGGCTTTGCG	ACCGGCGGCA	AAGCGCGCGG	GTTCGTGCGG	ACTTCGGGCG	480
ACAAATGGAA	GAACTGCGCA	TGCAAGACCG	TCACCGTCAG	GAATAAGGCG	AAGACCTACC	540
GGTGGACGTT	TGCGGACGTT	GAAGGCAAGG	TGCGGAGGAT	TAGGCTGATA	GACACCGGAA	600
GAGGCTCTGA	GGGCTGGGAA	TGTTAAAGCG	ATATGAGCGT	GTGCAAGGAT	GTGCTCTCTG	660
AGTTGAAGCT	ATGCGGCTAG	AGATTAAGCA	ATGAGGAGCG	AGATGCTCTG	GTGAGATCTT	720
CTTGAAGGAA	TGAAGAGG					739

3. INFORMATION FOR SEQ ID NO:265:

(1) SEQUENCE CHARACTERISTICS:

A LENGTH: 63 base pairs
B TYPE: nucleic acid
C STRANDEDNESS: single
D TOPOLOGY: linear

AGACGTCGTC GAGGCCGCCA TCGCCCGCGC CGAAGCCGTT AACCCGGCAC TGAACCGCTT 60
GGCGTATGC 59

(2) INFORMATION FOR SEQ ID NO:266:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:266:

ACTGCACCCG GCAGGCCGCA CCAACCGATC GGTCAACTA GCACTGCCCG TGGAGGCCCC 60
CCCGCGGTCT GTGCTTCCG ACCGGGAACG CTGGGGCAGC GCUGCTCCAG AAGGGTTGGA 120
GGGAGAGTTG GACBACCGTA TCGACGAGCG GTTCCCGGTC TTCAGCTCGG CCAGTCTCGC 180
CGAAGCCGCT CCGGGTCCCG TGACCCCGAT GACGCTGGAT CTCGAGTTGA CTCGACTGCC 240
CGCGGCCCGT CGGGCGATCG CTCGGCTACT GGTGCTTGGC GGTGTCTTG CCGATGAGTG 300
GGAGAGAAGA GCCATCGCGG TGTTCGGTCA CCGCCCGTAT ATCGGAGTGT CGGCCAATAT 360
TGTGCGCGCC GCCCAACTGC CCGGGTGGGA CGCGCAGGCC GTAACCCGGC GGGCACTGGG 420
CGAGCAACCG CAGGTCACCT AGCTGCTTCC GTTGGTCTGA CCGCAACTTG CGGGCGGACC 480
GCTCGGCTCG GTCGCGAAGG TGGTCTGAC GGCACGGTCG CTG 523

(2) INFORMATION FOR SEQ ID NO:267:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:267:

TTTCTGGTGT CTTGGGGTA GGAGCGACTT CCGCCCGCGG CGTGGCGCGT GGAGCGGGCT 60
TTGAGGAGAC CGGTGCCCGG TCGGCGGCGT GGGGACCAA AGGCGCGATC TATTGGGCGA 120
TTTCTGGTGT CCGGCGGCTT TCGGCGGCGT TTTGGGCTT CCGGCGGATC TTTTGGGCGA 180
TTTCTGGTGT CCGGCGGCGT TCGGCGGCGT TTTGGGCTT CCGGCGGATC TTTTGGGCGA 240

(2) INFORMATION FOR SEQ ID NO:268:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

```

TGAAGTGGTC TTGGCCATAG GGGGCGCTTG TCTCGGCGCTT GTCCAAACCG TGCAGCGCGC 120
CGTAGTGGCG TTGGTTGAGC CGCAGGTAC GCGGCACGGG AATCCAGAGC CGATCGGCGC 180
TGTCCAACGC CAGATGCGCG STGGTGATCG CGC3CCGCG CACGAGGTG TAGAGCACGT 240
CGGGCAATAG GTCGTGTTCC GCGATCAGCT CGCGGCTTCC AACC6CCTCT GCSTGGCCCT 300
TGTCGCTCAG GCGGACATCG ACCCAGCCCG TGAACAGGTT GAGGGCATTG CAGTCGCTCT 360
CGCCGTGGCG CAGCAACACC AGGCTGCGAG TGTTTGCCAT ACCGCAAGT CTCTCACGCA 420
CTCCCGCACT CCTCATCGTG GACCAAAATG CCGGAATTCT CCTCGGTCCG CTGCGCAGCG 480
CGTTCATACC GCGGAGG1GG TCGGCACCGT AACGGCCCGT T 521

```

(2) INFORMATION FOR SEQ ID NO:269:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:269:

```

GTCCAGGCTC ATTGGTCTGA ACAAAGCCAC CGGGCGCTAC AGCGGACGCG CCCATTGCTT 60
CTCGTGTATAG TCGCGGTACA GCTGGGCATC GGGGCGCTGGA CGAACCTCGG CCCAGG3GCA 120
CGGAACGAGC CGGTG3GCG TCACGCGGGG TCAGAACGGT AGTGACAGAC AGTCTCGCGG 180
CGCGAAGGGT TTGACGCGTC AGACTCGGCG TCGGCGTCTT CGACCGAGGC GTGGATCGCG 240
CGAGCTGAG AGCGTAGCGC CTCGAGCTCA CGGCGGAGCG GTTCCAGCAC CCACTCCACC 300
TCGCTGCTCT TGTTCGCGCG CAGCACCTGC GTGAACCTGA CGCGGTGAG ATCGGCGCGG 360
GTGACCGCGA AC3CGGCGAG CGTCTCGCGG TCGCTCGCGG GCGGCGAGGG CGGCAACTGC 420
TCGCGA 426

```

(2) INFORMATION FOR SEQ ID NO:270:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 118 base pairs
- B TYPE: nuc eic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:270:

```

TGGGAAGCG TGGGAAGAGT GCATGCGCT TGGGCGCTG TGGGCGCTG TGGGCGCTG 60
AGGCGCGCGC CGGCGCGCGA GTGTGGGCTA GTGCGCGTGC TGGGCGCTG TGGGCGCTG 120
TGGGCGCTA TGGGCGCGCG AGGCGTGGCT GTTCAAGCGA GCATGCGCT TGGGCGCTG 180
TGGGCGCTG TGTGCGCGAA CGGCGCGCGT TGGGCGCTG

```

(2) INFORMATION FOR SEQ ID NO:271:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 101 base pairs

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

AAGATCATCG	GCGCCCTCC	TTAGCATCGC	TGCGCTCTGC	ATCGTCGCGG	GCGCGGATCA	60
CGGAGGTCCG	GCCTTGATCC	CCACTCCTCG	AACGGTCAGC	ACCACAGTCG	GGTTCTCGGG	120
ATCCCTTTTCG	ACCTTGCCCC	GCAGACGCTG	GACATGCACG	TTCACCAGCC	TGGTATCGGC	180
TGGGTGCCCG	TAACCCCATATA	CCTGTTCGAG	CAGCACATCA	CGAGTAAACA	CCTGGCGCGG	240
CTTGCGCGCC	AATGCGACCA	ACAGGTCGAA	TTCCAGCGGT	GTCAACGAGA	TCTGCTCACC	300
GTTCGAGTG	ACCTTGTCG	CCGGTACGTC	GATTTCTACG	TCGGCGATGG	ACAGCATCTC	360
GCGGGGTTTCG	TCGTGCTTGC	GCGCGAGCCG	CGCCCGCACC	CGCGCAACCA	GCTCCTTGGG	420
CTTGAACGGC	TTTATGATGT	AGTCGTCGCG	GCCCCGACTCC	AGACCCAGCA	CCACATCCAC	480
GGTGTGCGTC	TTTGCGGTGA	GCATCAGGAT	CGGAACACCG	GAATCGGCGC	GCAACACCCG	540
GCACACGTCG	ATGCGGTTCA	TACCGGGGCA	A			571

(i) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 93 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

Leu	Phe	Gly	Ala	Gly	Gly	Val	Gly	Gly	Val	Gly	Gly	Asp	Gly	Val	Ala	1	5	10	15
Phe	Leu	Gly	Thr	Ala	Pro	Gly	Gly	Pro	Gly	Gly	Ala	Gly	Gly	Ala	Gly	20	25	30	
Gly	Leu	Phe	Ser	Val	Gly	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Ile	Gly	Leu	35	40	45	
Ala	Gly	Asn	Ser	Gly	Ala	Gly	Gly	Ser	Gly	Gly	Ser	Ala	Leu	Leu	Trp	50	55	60	
Ser	Asp	Gly	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Val	Gly	Ser	Thr	Thr	Gly	65	70	75	80
Gly	Ala	Gly	Gly	Ala	Gly	Gly	Asn	Ala	Ser	Leu	Leu	Val				85	90	95	

(i) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 93 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

20

25

(2) INFORMATION FOR SEQ ID NO:274:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

Lys Pro Asp Arg Pro Ala Ala Thr Val Gly Ser Cys Thr Thr Val Arg
 1 5 10 15
 Ala Pro Cys Ser Gln Pro Val Thr Thr Ala
 20 25

(2) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

Pro Pro Ala Gly Arg Pro Met His Pro Ala Pro Gly Thr Ser Ala Asp
 1 5 10 15
 His Pro Pro Asn
 20

(2) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

Ala Leu Val Ala Gly Cys Ser Ser Asn Pro Leu Ala Asn Thr Ala Pro
 1 5 10 15
 Thr Ser Pro Gly Thr Ser Ser Ala Ser Thr Val Ser Ser Thr
 20

```

      50              55              60
Pro Gly Ala Asp Ser Ala Ala Pro Ala Ser Ile Met Val Phe Asp Asp
65              70              75              80
Met His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala
      85              90              95
Leu Thr Ser Asp Asp His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
      100              105              110
Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
      115              120              125
Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala
      130              135              140

```

(x) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

```

Met His Ile Thr Leu Asn Ala Ile Leu Arg Ala Ile Phe Gly Ala Gly
1              5              10              15
Gly Ser Glu Leu Asp Glu Leu Arg Arg Leu Ile Pro Pro Trp Val Thr
      20              25              30
Leu Gly Ser Arg Leu Ala Ala Leu Pro Lys Pro Lys Arg Asp Tyr Gly
      35              40              45
Arg Leu Ser Pro Trp Gly Arg Leu Ala Glu Trp Arg Arg Ala Tyr Asp
      50              55              60
Thr Val Ile Asp Glu Leu Ile Glu Ala Glu Arg Ala Asp Pro Asn Phe
      65              70              75
Ala Asp Arg Thr Asp Val Leu Ala Leu Met Leu Arg Ser Thr Tyr Asp
      80              85              90              95
Asp Gly Ser Ile Met Ser Arg Lys Asp Ile Gly Asp Glu Leu Leu Thr
      100              105              110
Leu Leu Ala Ala Gly His Glu Thr Thr Ala Ala Thr Trp Ala Gly Arg
      115              120              125
Met Asn Gly Ser Thr Ile Thr Arg Thr Thr Thr Arg Leu Thr
      130              135              140

```

(x) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 161 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

Val Leu Val Ala Gly Cys Ser Ser Asn Pro Leu Ala Asn Phe Ala Pro
 1           5           10           15
Gly Tyr Pro Pro Thr Ile Glu Pro Ala Gln Pro Ala Val Ser Pro Pro
          20           25           30
Thr Ser Gln Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro
          35           40           45
Arg Ala Ala Leu Phe Asp Asn Gly Thr Arg Gln Leu Val Ala Leu Arg
          50           55           60
Pro Gly Ala Asp Ser Ala Ala Pro Ala Ser Ile Met Val Phe Asp Asp
65           70           75           80
Val His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala
          85           90           95
Leu Thr Ser Asp Asp His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
          100          105          110
Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
          115          120          125
Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala Arg Arg Ser Asp
          130          135          140
Gly Lys Leu Val Leu Gly Ser Ala Asp Gly Ala Val Tyr Thr Leu Ala
145          150          155          160
Lys Asn Pro

```

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:279:

```

Trp Gly Ala Pro Pro Ser Gly Gly Pro Ser Pro Trp Ala Gln Thr Pro
 1           5           10           15
Arg Lys Thr Asn Pro Trp Pro Leu Val Ala Gly Ala Ala Val Val
          20           25           30
Leu Val Leu Val Leu Gly Ala Ile Gly Ile Trp Ile Ala Ile Arg Pro
          35           40           45
Leu His Val Gln Ser Ser Asn Pro Val Val Ala Gly Arg Leu Ser Val
          50           55           60
Ser Leu Leu Asp Ser Ser Gly Val Asn Ala Val Met Gly Ser Ser Ser
          65           70           75           80
Met Gln Pro Gly Leu Pro Ile Thr Ser Met Asp Ser Ser Pro Val Thr
          85           90           95
Val Ser Leu Pro Asp Lys Gln Gly Ala Leu Tyr Thr Ser Gln Asp Pro
          100          105          110
Val Thr Ala Gly Thr Thr Thr Thr Ala Ile Asn Gly Leu Ile Ser Ser
          115          120          125

```

```

Lys Trp Lys Asn Cys Ala Gly Lys Thr Val Thr Val Thr Asn Lys Ala
      165      170      175
Lys Thr Tyr Arg Trp Thr Phe Ala Asp Val Lys Gly Ser Pro Pro Thr
      180      185      190
Ile Thr Val Ile Asp Thr Gln Glu Gly Ala Glu Gly Trp Glu Cys Gln
      195      200      205
Arg Ala Met Ser Val Ala Asn Asn Val Val Val Asp Val Asn Ala Cys
      210      215      220
Gly Tyr Gln Ile Thr Asn Gln Ala Gly Gln Ile Ala Ala Lys Ile Cys
      225      230      235      240

```

(2) INFORMATION FOR SEQ ID NO:280:

1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

11. MOLECULE TYPE: protein

xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

```

Asp Val Val Glu Ala Ala Ile Ala Arg Ala Glu Ala Val Asn Pro Ala
 1          5          10          15
Leu Asn Ala Leu Ala Tyr
      20

```

2) INFORMATION FOR SEQ ID NO:281:

1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

11. MOLECULE TYPE: protein

xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

```

Leu Ala Pro Ala Ile Ala Thr Asn Gly Ser Gly Ile Leu Ala Leu Pro
 1          5          10          15
Ala Ile Ala Pro Pro Arg Ser Val Pro Ser Ala Gly Ile Pro Leu Gly
 16          20          25          30
Ser Ala Ala Pro Ile Gly Leu Ala Ile Ile Thr Asn Asp Arg Ile Asp
 31          35          40          45
Ile Arg Phe Pro Val Phe Ser Ser Ala Ser Leu Ala Glu Ala Leu Pro
 46          50          55          60
Gly Pro Leu Thr Pro Met Thr Leu Asp Val Ile Leu Ser Gly Leu Arg
 61          65          70          75          80
Ala Ala Phe Ala Ala Met Thr Thr Leu Ala Leu Gly Gly Val Val
 81          85          90          95

```

```

      115              120              125
Trp Asp Ala Gln Ala Val Thr Arg Arg Ala Leu Gly Glu Gln Pro Gln
      130              135              140
Val Thr Glu Leu Leu Pro Phe Gly Arg Pro Gln Leu Ala Gly Gly Pro
      145              150              155              160
Leu Gly Ser Val Ala Lys Val Val Val Thr Ala Arg Ser Leu
      165              170

```

(2) INFORMATION FOR SEQ ID NO:282:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

ii. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:282:

```

Val Gly Val Val Gly Val Gly Ala Thr Ser Pro Ala Gly Ala Gly Ala
 1              5              10              15
Gly Ala Gly Ser Ala Gly Thr Gly Ala Gly Ala Gly Gly Gly Ala Thr
      20              25              30
Lys Gly Arg Ile Asp Ser Ala Ser Ala Leu Ala Ala Pro Leu Ser Thr
      35              40              45
Gly Leu Leu Ala Val Pro Ser His Thr Thr Asn Gln Arg
      50              55              60

```

(2) INFORMATION FOR SEQ ID NO:283:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 133 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

ii. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:283:

```

Val Ala Asn Thr Ala Ser Leu Val Leu Leu Arg His Thr Ile Ser Asn
 1              5              10              15
Leu Asn Ala Ser Asn Ser Phe Thr Gly Trp Val Asn Val Ile Leu Thr
      20              25              30
Asn Lys Gly Gln Ala Gly Ala Val Arg Ser Gly Gln Leu Ile Ala Ile
      35              40              45
His Asp Leu Leu Pro Asp Val Leu Tyr Thr Ser Leu Leu Arg Arg Ala
      50              55              60
Thr Thr Thr Ala Val Leu Ala Leu Asp Ser Ala Asn Arg Leu Trp Ile
      65              70              75

```

Phe Met Ala Trp Arg Arg Ser Tyr Asp Thr Pro Pro Pro Pro Ile Glu
 115 120 125
 Arg Gly Ser Gln Phe
 130

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:284:

Pro Gly Ser Phe Ala Arg Thr Cys Pro Pro Gly Arg Thr Ala Asp Ala
 1 5 10 15
 Pro Ile Arg Cys Arg Asp Ser Arg Gly Thr Ala Gly His Arg Ala Leu
 20 25 30
 Asp Glu Pro Pro Pro Arg Gly Ser Glu Pro Ala Arg Arg Arg Ser Arg
 35 40 45
 Gly Val Arg Thr Val Val His Asp Ser Leu Ala Ala Arg Arg Val
 50 55 60

(2) INFORMATION FOR SEQ ID NO:285:

(i) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 73 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

Gly His Gly Gly Gln Ser Ala Ile Gly Leu Gly Gly Gly Ala Gly Gly
 1 5 10 15
 Asp Gly Gly Gln Gly Gly Ala Ile Arg Leu Leu Trp Gly Thr Gly Gly
 20 25 30
 Gly Gly Gly His Gly Gly Ala Asp Arg Thr Tyr Arg Ile Thr Ala
 35 40 45
 Ala Arg Ser Gly Arg His Gly Arg Arg Gly Trp Arg Arg Trp Ala Asp
 50 55 60
 Arg Gln Arg Arg Gly Arg Arg Arg
 65 70

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

```

Asp His Arg Arg Arg Ser Leu Ala Ser Leu Arg Ser Ala Ser Ser Pro
 1           5           10           15
Ala Arg Ile Thr Glu Val Arg Pro Cys Thr Pro Leu Leu Glu Arg Ser
 20           25           30
Ala Pro Gln Ser Gly Ser Arg Asp Pro Phe Arg Pro Trp Pro Ala Asp
 35           40           45
Ala Gly His Ala Arg Ser Pro Ala Trp Tyr Arg Leu Gly Ala Gly Asn
 50           55           60
Pro Ile Pro Val Arg Ala Ala His His Glu
 65           70

```

2. INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

```

CCGCGACGTAA CACCCTGAAT TGAAGGGAGG CCGTGGTCAT GGGCCGATTC TATCCGTGGG      60
CGAACGGTTA TTGACGGGCC GGAGGGGACT CCGCTGCGAC CAGTGGTCA CTCAGCGCGT      120
TTTCNCGGCA ACCAAGCGGC GACATACCA TTGACATTCC ACAGCACGGC CCGC      174

```

2. INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

```

TCCCAACGCG GTTGACGTTG CCGCGGCTGG CCGTAAAGAG TTCTCGGCAC TTTCGGGTGA      60
CCCTCGCCCG GCACCGGAGG ACCGCTGAGC TCAAGCTGCT AGTGGAGGTG CTCGACGGTG      120
CCCTGGGCAC GATGGGCTCT GAAAGCGTGG GCAGCGGGGT GCTGGCTGTG TTAAAGCGCT      180
CGCTCAGCGG CCGGATCTGG CCGCTGCGCG ACCTCGACAT TCTGCTGGAG GGTGAGCAAG      240
ATCGACCGCG ACCGCGCTGG CCGGATCTCA CCACTCGCGG ACCGCGGGTG TATAGCGGCT      300
TCCCGGAAAT TCGTGGGCGG CAGCTGACT CCGTGGCTCT CAGTGGGCGG CAGTGGGCGG      360

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

Ala Asn Gly Val Thr Phe Arg Pro Val Ala Leu Glu Ser Leu Ser His
 1 5 10 15
 Phe Pro Val Thr Val Ala Ala His Arg Ser Thr Gly Glu Leu Thr Leu
 20 25 30
 Leu Val Glu Val Leu Asp Gly Ala Leu Gly Thr Met Ala Pro Glu Ser
 35 40 45
 Leu Gly Arg Arg Val Leu Ala Val Leu Gln Arg Leu Val Ser Arg Trp
 50 55 60
 Asp Arg Pro Leu Arg Asp Val Asp Ile Leu Leu Asp Gly Glu His Asp
 65 70 75 80
 Pro Thr Ala Pro Gly Leu Pro Asp Val Thr Thr Ser Ala Pro Ala Val
 85 90 95
 His Thr Arg Phe Ala Glu Ile Ala Ala Ala Gln Pro Asp Ser Val Ala
 100 105 110
 Val Ser Trp Ala Asp Gly Gln Leu Thr Tyr Arg Glu Leu Asp Ala Leu
 115 120 125
 Ala Asp Arg Leu Ala Thr
 130

2 INFORMATION FOR SEQ ID NO:290:

1. SEQUENCE CHARACTERISTICS:

- A LENGTH: 326 base pairs
 B TYPE: nucleic acid
 C STRANDEDNESS: single
 D TOPOLOGY: linear

11. MOLECULE TYPE: cDNA

11. SEQUENCE DESCRIPTION: SEQ ID NO:290:

ATTGTAAGG	TTAGCAATAC	CTGTTCTGGG	TGGTTTGTGG	AGGTCCTAT	GAGGAAAGG	91
TTAAGAAGA	TACCAAGGCG	ATGCGCGACG	TGTTGCGGGT	TGCGGCGGTT	AAATATTTG	121
CTGTGGGGT	TGGGAAAGG	TGCACCGGCG	ACTCGCGCGG	AGCGTCGCGG	AACGAGTTTG	130
CTGTCTAGCA	GTGGGCAGAA	CAGCGCGCTG	AGACCGCTGA	CGGTTGCTTG	GAGGCTGTGG	240
AGACCTGTGA	TGCAAGATG	GTGTCACCT	GCGCGCAGTG	CTTCAACAGC	ATGCGCAAGG	300
AAATATCGGA	CTGCGCGCGG	AACTACACCG	TGCTGCACCA	CAGCGAGCTG	CTCAATCGGT	360
TGTTGCGCGA	CAAGAGGCTG	GTGCGTGTCA	CTCGCGTTTG	TCAGGACATG	ACGTACCAAG	420
ACCGCTGCTA	CTGCGGTGGG	GAGAACAAAG	TCTACGAGGC	ACCACGGGAG	CTGATCGGTT	480
CTGCGCGGGT	CACTGATGCG	GAGATGCGCG	GCGATGCGGA	TCGCGAG		526

INFORMATION FOR SEQ ID NO:291:

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CTCGCCGCCG	TGATCTGGCC	GGCGAACTTC	GTCAGTGCAT	CCAGACCCCA	ACGATCATCG	60
ATCAGGCCGA	TGCCCATGAT	CACCGCACCG	GCCACCAGCA	CCGCGGGCAT	GCCGGTGGAA	120
TAGACGAACC	CCCGGGTGAG	TGCCGGAAGC	TGGGAGGCAA	GAAAGACGGC	GCCGACAATG	180
CCCAGGAACA	TGCGCAACCC	ACCCATCCGA	GGGGTAGGCG	TGACGTGCAC	ATCTCGCTCC	240
CCCGGGTAGG	CGACGGCTCC	CAGGCGACTG	GCCAGCATCC	GCACCGGACC	GGTCGCAAAA	300
TAGGTGATGA	TGCGCCCGGT	CAGCCCGACC	AGCGCAAGCT	CACGCAGCGG	GACACCGCGG	360
CCGCGATAGG	ACAGGGCGAG	CAAGCCACCG	GCAACGCCGG	CCACATCGCT	GGACACCTCG	420
AGACCGTACT	GCACCAACCT	GAAGAGCTGA	ACAATCGCCG	AACGTGCAAC	AGCTGCCAAC	480
AATTGGG						497

2. INFORMATION FOR SEQ ID NO:292:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

ATGAAGCCCG	ACAATATGAG	CCGGGGCAAC	CCGGCATGTA	CGAGCTTGAG	TTCCCGGCCG	60
CTCAGCTCTC	CTCCTCCGAC	CGCCGTGGTC	CGTGTGTGGT	GCACGCTTTC	GAAGGTTTCT	120
TCGAAGCCCG	CCATCCGATC	CGGCTGGCCG	CCGCCACCT	CAAGGCGGCG	CTGGACACAG	180
AGCTGTGCGC	CTCCTTCCCG	ATCGATGAAC	TACTGGACTA	CGGCTCGCGG	CGGCCATTAA	240
TGACTTTGAA	GACCGATCAT	TTGAGCCACT	CGCATGATCC	TGAGCTAAGC	CTGTATCCGC	300
TGCGCGACAG	CATCGGCACC	CCATTCTCTC	TGCTGSCGGG	TTTGGAGGCG	GACCTGAAGT	360
TGGAGCCGGT	CATCACCGCC	CTCGGATTGC	TGCCCJAGCG	CGTGGGTGTA	CGGCAGAACC	420
ATCGGCGCTG	GCACCGTCCC	GATGGCCGTT	CGGCACACAC	GACCGATCAC	GATGACCGCT	480
CATTGCAACA	ACCGGGAGCT	ATCTCGGATT	TTCAACCGTT	CGATCTCC		528

3. INFORMATION FOR SEQ ID NO:293:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 512 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

```

GGCCGTTGCC GCGGCTGGTG GGGGTGGCGG CCTGGTTGAC GTATTGTTCC ACCGGCCCCGG 240
CCCTTGACCC TTGGGCGGTG TCGATCGCGG CGTCGATGGA TCCGCCGACC ACGACGTGCG 300
AAGCCTCGCC TGCCGCGGCA GCGGCCCAAC TGTGTCGCGG CTCCTGCGAT TTGGCCCCGG 360
CCGACGAGAT GATGGGCACC ACCGAGGCCT GCGGCCGTCT GGGGGAGGCC AGCGCGGGTT 420
CGCGGTCACG CCATACGCGA CGGTGCGCGG CCGCTTCGGA GATTTCAGG CTGCGTTGCA 480
CCAGATCGAG CAGCGGTCTG CCCAGGGACT GGGTTAGCCC GTTGGCGCCG CCGTTGTAGC 540
GGCGAGCGCA ATATCGGTGC CCACTCGACC CAACCGCGAC TCCATAAGCG ACACCATTCG 600
CGGTCGATGC

```

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 164 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

```

Phe Asp Gly Tyr Glu Tyr Leu Phe Trp Val Gly Cys Ala Gly Ala Tyr
1      5      10      15
Asp Asp Lys Ala Lys Lys Thr Thr Lys Ala Val Ala Glu Leu Phe Ala
20     25     30
Val Ala Gly Val Lys Tyr Leu Val Leu Gly Ala Gly Glu Thr Cys Asn
35     40     45
Gly Asp Ser Ala Arg Arg Ser Gly Asn Glu Phe Leu Phe Gln Gln Leu
50     55     60
Ala Gln Gln Ala Val Glu Thr Leu Asp Gly Leu Phe Gln Gly Val Glu
65     70     75     80
Thr Val Asp Arg Lys Ile Val Val Thr Cys Pro His Cys Phe Asn Thr
85     90     95
Leu Ile Lys Glu Tyr Arg Gln Leu Gly Ala Asn Tyr Thr Val Leu His
100    105    110
His Thr Gln Leu Leu Asn Arg Leu Val Arg Asp Lys Arg Leu Val Pro
115    120    125
Val Thr Pro Val Ser Gln Asp Ile Thr Tyr His Asp Pro Tyr Tyr Leu
130    135    140
His Arg His Asn Lys Val Tyr Glu Ala Thr Arg His Leu Ile Glu Ala
145    150    155    160
Val Gly Ala Thr

```

(2) INFORMATION FOR SEQ ID NO:295:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 161 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

Arg Arg Arg Asp Leu Ala Gly Glu Leu Arg Gln Cys Ile Gln Thr Pro
 1 5 10 15
 Thr Ile Ile Asp Gln Ala Asp Ala His Asp His Arg Thr Gly His Gln
 20 25 30
 His Arg Gly His Ala Gly Gly Ile Asp Glu Pro Pro Gly Glu Cys Arg
 35 40 45
 Lys Leu Gly Gly Lys Lys Asp Gly Ala Asp Asn Ala Gln Glu His Arg
 50 55 60
 Gln Pro Thr His Pro Arg Gly Arg Arg Asp Val His Ile Ser Leu Pro
 65 70 75 80
 Arg Val Gly Asp Gly Ser Gln Ala Thr Gly Gln His Pro His Arg Thr
 85 90 95
 Gly Arg Lys Ile Gly Asp Asp Arg Arg Gly Gln Pro Asp Gln Arg Lys
 100 105 110
 Leu Thr Gln Arg Asp Thr Gly Ala Ala Ile Gly Gln Gly Gln Ala
 115 120 125
 Thr Gly Asn Ala Gly His Ile Ala Gly His Leu Glu Thr Val Leu His
 130 135 140
 Gln Pro Glu Glu Leu Asn Thr Arg Arg Thr Cys Asn Ser Cys Glu Gln
 145 150 155 160
 Leu

12. INFORMATION FOR SEQ ID NO:296:

12.1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

12.2 MOLECULE TYPE: protein

12.3 SEQUENCE DESCRIPTION: SEQ ID NO:296:

His Ala Arg Glu Tyr Glu Pro Gly Gln Pro Gly Met Tyr Glu Leu Glu
 1 5 10 15
 Phe Pro Ala Pro Gln Leu Ser Ser Ser Asp Gly Arg Gly Pro Val Leu
 20 25 30
 Thr His Ala Leu Glu His Phe Ser Asn Ala Gly His Ala Ile Arg Leu
 35 40 45
 Ala Ala Ala His Leu Lys Ala Ala Leu Asp Thr His Leu Val Ala Ser
 50 55 60
 Phe Ala Ile Asp Glu Leu Leu Asp Tyr Arg Ser Arg Arg Pro Leu Met
 65 70 75 80
 Thr Phe Lys Thr Asp His Phe Thr His Ser Asp Asp Pro Glu Leu Ser
 85 90 95
 Leu Tyr Ala Leu Arg Asp Ser Ile Gly Thr Phe Phe Leu Leu Ala
 100 105 110
 Leu Leu Glu Phe Glu Leu Leu Thr Glu Arg Phe Ile Thr Ala Ala Arg
 115 120 125

145 150 155 160
 Phe Gln Gln Pro Gly Ala Ile Ser Asp Phe Gln Pro Phe Asp Leu
 165 170 175

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

Lys Pro Val Lys Glu Pro Val Pro Ala Leu Pro Pro Val Pro Pro Thr
 1 5 10 15
 Pro Ala Leu Pro Pro Leu Pro Pro Leu Pro Pro Val Pro Gly Phe Pro
 20 25 30
 Thr Val Pro Pro Pro Gly Ser Met Ala Pro Leu Phe Arg Pro Phe Ser
 35 40 45
 Pro Ala Pro Pro Ser Pro Ala Leu Pro Pro Ser Pro Pro Leu Pro Pro
 50 55 60
 Leu Val Gly Val Ala Ala Trp Leu Thr Tyr Cys Ser Thr Gly Pro Ala
 65 70 75 80
 Leu Asp Pro Leu Ala Val Ser Ile Ala Ala Ser Met Asp Pro Pro Thr
 85 90 95
 Thr Thr Cys Glu Ala Ser Pro Ala Ala Ala Ala Ala Gln Leu Cys Arg
 100 105 110
 Gly Ser Cys Asp Leu Ala Pro Ala Asp Glu Met Met Gly Thr Thr Gly
 115 120 125
 Ala Cys Gly Arg Leu Gly Glu Ala Ser Ala Gly Ser Arg Ser Arg His
 130 135 140
 Thr Arg Arg Cys Ala Ala Ala Ser Glu Ile Cys Arg Leu Arg Cys Thr
 145 150 155 160
 Arg Ser Ser Ser Gly Val Pro Arg Asp Trp Val Ser Pro Leu Ala Pro
 165 170 175
 Pro Leu

(3) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 221 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

```

ATTTCACCCC  ANGCAGCTAC  CACACGGGGA  CTCGGGAAACA  CCGGCGATTT  TACACCGGCS  180
CCTTCATCTC  CGGCAGCTAC  AGCAACGGGT  TTTGTGGAGT  GGAAATTATC  AGGGCTCATT  240
GGNTGCACCC  GGSCTTRCGA  ATCCCTCGKG  CCAATTCAAC  TCCTCNACAA  GCTTGCGGCC  300
GCACCTCSAGC  CCGGGTGAAT  GATTGAGTTT  AACCCTNAN  CAATAACTAG  CATAACCCCT  360
TKGGGCGCTT  AAACGGGTCT  TGAAGGGTTT  TTTGCTGAAA  GGANGAACTA  TATCCGGATA  420
ACTGGCGTAN  TACGAAAAGC  CGCACCGATC  GCCTTCCCAA  CAGTTGCGCA  CCKGAATGGC  480
AATGGACCNC  CCTKTTACCG  GSCATTAACN  CGGGGGTGTN  GGKGTTACCC  CCACGTNACC  540
GCTACCTTGC  CANNSSCCTN  RSGCCGTCTT  CGTTTCTTC  CTTCTTCTC  CCMCTTCGCC  600
GGTTCCCTTC  AGCTCTAAAT  CGGGGNNCCC  TTTMGGGTTT  CAATTATTGC  TTACNGSCCC  660
CCACCCCAAA  AAYTNATTNG  GGTAAATGTC  CCTTMTTGGG  CNTCCCCCTA  WTNANNGTTT  720
TCCCCCTTNA  CTTTGRSTCC  CTTCTTATW  NTGAMNCTNT  TTCCACYGGA  AAAMNCTCCA  780
CCNTTYSSTS  TTTCTTTTGA  WTTATMRGGR  AATTSCAATV  CCGCVTTKGG  TTMAANTTAA  840
CYTATTTTNA  ATTTTCCCGM  TTTTMMNATR  TTNSNCKCGM  FNCTCCNRKA  SSGNTTTCCT  900
CCCCCYTTSS  GKTVCSCCRN  G

```

921

(2) INFORMATION FOR SEQ ID NO:299:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11) MOLECULE TYPE: Genomic DNA

X1) SEQUENCE DESCRIPTION: SEQ ID NO:299:

```

AATTCGGCAC  GAGATANGGG  CCGACCCGGG  TCGGCAAGCG  GCGGACCCCT  GCGCAGCATC  60
ACCGGGGTCA  ACAGCACCAC  GGTGGCGTCC  ANGCAGAGCG  CCGCAGTGAT  GCGGCGCGAG  120
ACGGGCAACA  CCGCCCGTAG  CAGTCGGTGC  GACTCGCGCG  TCGCTCGANC  CATGGCCCGG  180
GUGGCTGCTT  CGAACANGCC  TCGTCTCTCC  ACAGCTTAGC  CAGCANCCAA  ACCGCACCCA  240
GAAACCCACA  CGCCCCCGCC  CCGGANACC  TCGGATATCG  KCTGCTGGGG  CGANATCCCG  300
CGATCCCTNA  CANGATGAGC  CTTGCCGGAA  CCGCGCCGCT  GCTCGCGGGC  AGCCGCGTGG  360
CGCGGGCAAC  CGCGAAGCCA  NGAACACGGC  AAGCAATATC  ANCGCAACAG  CAATTGTCAA  420
GGGTAAACG  TTTCACATCC  AGGATCTCC  CCGCGCCACA  CGCTGGMTG  TGCAGSGCGA  480
TCCCTTCCTN  GGGCGGNCAC  TCTCAAAGA  TCGNGATCNA  CAGKCTAGGT  CTTGGGCGCA  540
TATGSAAGGN  CGCAACGGNT  TTAAAGCGGC  GAAAAAATC  TCCANTGGA  TAAATCAGC  600
CGGGGANCCT  CGCGTGSCMM  NGTCVCGGKC  ATTCTTCAAC  MGGTTTNAGC  GCGGKTGCNG  660
CGTAACTKGC  TAAAMTTAAG  KTNGGGNTY  CCGGCGCGTA  ACCGGCCTK  NGCCGCTTAA  720
AAAACCGGNC  TTTCTTKGAT  TMMACCGCN  TCGGAWTGG  CGGTGKTCC  CANGNTYAAC  780
AMGCTCCGCG  KNGGCTTGGG  GAAACCTTTC  TONGGCGTTC  NTGCTTCTTC  AWMGCGCGCG  840
AAACCGGKVC  TGTCTGCTTC  WASJAMNCC  MNNGTCTCT  TAAAGGSCAN  CNPAAWGKVT  900
TCTTGGGAAG  TCTTCAATTC  TAAAYTCTC  TTTMMGSCCN  TTTTCWRTTN  NRNGGGAAC  960
AMWTHYCTNC  TWTCAATTC  TTTTCCASMN  TAAAGCTTTC  TTTTCGSSC  STCCMGGSNC  1020
GGCTNNANAN  AAASATCTT  TTTNNNANKF  TTTTCTCTTC  TTTMGRENRE  BMGAACCCGR  1080

```

1082

2) INFORMATION FOR SEQ ID NO:300:

1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 1082 base pairs
 B TYPE: nucleic acid

11 MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:300:

```

AATTGGCAGC AGTGATCGCG CTGAAGCCCG TAGCGCGGGT GGCTCGGGTG GTTTGCCAAC    60
RAAATCCGCT CGANGTGGTC TCGGTAGCGG GTGTCCANAA CGGTGGCGCG GTGCCGGCGG    120
ATCTGATCGG CGCGGCCGTA GTGCACGTCG GCGGGCGTGT GCAGTCCGAT GCCGGAATGC    180
TTGTGTTTCT GGTGTACCA GCCGAAGAAC CGGTGCGAGT GCACCCGGGC CGCCTCGATC    240
GACTCGAACC GTTTCGGGAA ATCGGGCCGG TACTTGAAGG TCTYGAACG GGCCTCAGAC    300
AACGGGTGTG CTTGCTGGTG TCGGGCGGTG AGTGCGACTT GGTGACACCG AAGTCGGCCA    360
NCANCAATGC CACCGGTTTG GAACTCATCC ACAACCCCGG TCCGCGTCMA GGTCACTTGT    420
NCGGCGCTAA TTNYTGGGC GGCAAGGTTT TCCCGAYCAN KCCGCTCGGC CAAAACTTCG    480
ANTCNCSCCA AGGCCNCCAT CCNCCCAAAC AMGTTACGGG ANAAAAATY CAAAGAYCAC    540
CYTCCGGKTN TTATANCTYC CCYTTTGSTY GGGCCCCCGN CYVTGKKNAT ACCCTNCCA    600
AWTCCCAACN CCKKCCAANA RCYKGGGGCC CCNCCAACC CGGKGGAACA WTAATTTAAA    660
CCCYAACMAW ACTWMMNACC CNNGGGSCCY AAMCGTYNR AGGTTTTSTC NAAAGAAASA    720
ANTCGGAAMC CGGNTSTACC AAAAASCCCH CCNWTCCCTC CRADATTGSC NCESAANKSA    780
AKGCCCCUNY TCGGCMWNNC CCGCGGKKKT KKGTTNCCCT WMRCWMWYTS GGCENASCCN    840
CKYYSMYCC CCGCTCCCCM CTCGCKTCC CCAMCCYANC MGGCCCCCTM GKKCCCKNT    900
YKCCCCCCCC AMMNNGGGG WGACCTTNGG CCGCMKRRGM TCCCNANTGA MCCTCWGNRA    960
MECYCCNRAR ANMCCSCNCC NGCNCRCCKN

```

(2) INFORMATION FOR SEQ ID NO:301:

A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 223 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:301:

```

AATTGGGTTG GCAACGCGGG CCTGTTCCGC AACGGCGGGG CCGGTGGTGC CGGTGGGGCT    60
ATGAGTGGCG CGGCGCGGCC GGGCGGTAC CCGGGGTGGT TTGGTCATGG CGGCGGTGGC    120
AGGTGGGTTG GTGTANGTGC GCGCGCGGCC AACGGTGCTA CGCCCGGTGA GGATCGGGGC    180
ATGAGTGTTC CGGAGTGGGA CRACCTCTCT CCGGCTCGTG CCG

```

A. INFORMATION FOR SEQ ID NO:302:

A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 415 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:302:

```

TAGGACAGGC GTTGGCCGCA CTCACGCGC CCATCCAGG CGTGACCGG CCCCCCTGAT 300
GGGAACCGG CCAACGGGG CCGGGGCAAC GGGGCCCCG GCRGGCACG CCGGTGGTTG 360
TTCGGCGCG GAAGGAACG CCGGTCCGG GTCANCRGG GGGCGGGCG AATGCCCC 420

```

(2) INFORMATION FOR SEQ ID NO:303:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1049 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:303:

```

AATTCGGCAC GAGGGGSCAG ATCCCATACA TCGCTCCCGG CAGACCCGGT GATACAGCA 60
GTTGGGACCA GCGGAGCGCA CAATACGGG TCTGGGTGT CCGCTTGAGC ACCACCCCT 120
TACCGGCCAC CAGCGCGCGG ACCGATCGG ACACCGTAAG CCGCATGGGG TAGTTCCAC 180
GGGAGATCAC CCCCACCAAG CCGTTGGTT GATAGCACAC CGTGGTCTTG CCTATCCGG 240
GCACCAGCGG CTGTGCCTTA CCGGGTTCA GCAGGTCCAC ACAGACTCGT GCSTTATAAT 300
TTCGGSTTCC GCGATCAGAT CGACAATTG CTCTTCCCGG GCGCATCGGG CCGTGGCCCG 360
CTCGGCTTGC AGGAAGTCCA TGAAGAATC GCGGTTCTCG ATNAACAGGT CGGATAGCG 420
GCGGATCACT GCAGGTCCGT CGATNACGG ACCTTCGCCA CTCGGTCTGC GCGCGCGAN 480
CTTCGGCGAA TGCTGCTTCC ACTTCGGCGG NCGTGGCAAC CGAATCNTAT CAGGGGTTGC 540
CGGTTAAAC TCGTCAATST NCGGTGCA AATCGGCAAC TTCTTATCCC GCGAGGTRCC 600
AAGSANNCAA ACCTCGGCAA GGTAGGMIT TCGCCGNCCT YCAAAAATNC GGTCTTTGGN 660
GMAATTTCCG CKCNATGKTG MCAAGGMITT KKAANAACG GGGTCYTCTN NTGNGKGGAK 720
GCAAMGGHT TTGGGMAAG GKMNCCAAH CTWAGGCTG KTKAANGGNW TTCCCCCGGG 780
GGAKKNGA ATYCCGNA NCCGCGGGG GNMCAATTC TCCCGMCTC CTCGGGAWTC 840
GCGGCTTCC GAAAAACCG GCGAAATMM TCTTTCGCG TCTGAGACG GTTTHARGA 900
MMSSAARNG NMCHCTGCG CHTKTGKT AAAAAGNAYW GCGMAAATT TTTATTTCC 960
TTCGCGGCGN TCGCTNTTT TCGMTWETH WNTNORMCG MMSNCKSNG KCGNRCNN 1020
TTCGCGGCGN AAWNTKCGN TTTATMAGC 1040

```

(2) INFORMATION FOR SEQ ID NO:304:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:304:

```

AATTCGGCAC GAGGGAATTC AGAATCCCGG AATGCTAAG CCGGGTCTG TCGGCTTACG 60
TGAAGATGCA GGTGAGGCTT TCGGGCTG AATGCTGCA GGTGAGGCTT TGAAGATGCA 120
ATGCTGCTG TGAAGATGCA GGTGAGGCTT TGAAGATGCA AATGCTGCTT TGAAGATGCA 180
TGAAGATGCA AATGCTGCTT TGAAGATGCA AATGCTGCTT TGAAGATGCA 240

```

YTTATTTTMS	GCTNAYGGGA	ATBAMRGGA	CAAYNTCCCT	CCCMGGAAAA	ACCAACMSGC	540
CCTGGTNSYC	CNCCCRCCNC	AKAACCCRTT	KCTGTRSTM	CCSMAAATNA	CSCCCSCTTS	600
NACTCCNCSC	AANTNSCCCC	CCCSCKNNTT	ATSTYCCCGK	GTTCCCGGMC	CCCTTNAAMC	660
TCCCCGGTTA	ACCCGCTWNT	SNCNCCCCCS	YTAACKMCRG	GCTTSTTNGT	CCCCCYTRMK	720
CNCCCCCTCK	SAMCWNCNC	CTCKAACNAC	CCCKCYKGS	TNCCCAATNT	WCMWCKCNS	780
KTNTMTCTK	CCAAYTNCRC	CCNCRCTCCC	CKSTSTCAM	WTATAAAACG	WCWYAWYNNK	840
KCNWMAWTA	MGACWCTCNY	NCCCNCCNCK	NTTKTAMWCC	CKMCCCKCSW	TWCYCKCSCC	900
CCMTCTMNAC	YCCCCCKKTY	NKWMCCCTTC	CCCCCTCCC	MCNMBMKTCT	YCSGKTWCWC	960
NCYNTTMTCN	CYNANMCKCK	KTCTCTTCN	CRNTCTCCCC	CCWCCCCCCT	KKCTCTSKCC	1020
CNCNCTCCSC	MMKGS					1036

(2) INFORMATION FOR SEQ ID NO:305:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(B) MOLECULE TYPE: Genomic DNA

(C) SEQUENCE DESCRIPTION: SEQ ID NO:305:

AATTCGGCAC	GAGATCATCA	ATAGCGGGCT	GGTCAGCACC	GAAGTGGTCC	CGGATCTCCG	60
GAGCAAGTCT	CGTCTGCTCG	CCCAGCAGGA	GGTCGGCACC	GATCGCGACA	CCTGCGATGT	120
TTTGGATGGT	GTTCAGTTCC	AGGTAAGGCC	GAAGCGGCAG	CTTTGCTAGC	AGGGTGTCTT	180
TGCTTTTCGC	ACGTGACGTA	ACCAATAACT	GGAGCGCAGA	CCAACTCCCG	CCCTCGATCC	240
GGGTACGAGG	GTCCGCCCCA	CCCAGCCGTT	GTCCGCGCTG	GGCCGAAGGT	CAGCTGCTGT	300
GGGATCGAAG	TAAGAAACCG	CGCCATGCCC	GTGGCAGGCT	ACGACTGACC	GAGCAAACGA	360
ACGATCGTCC	TGCTTTCCCT	GGGGGTAATC	GAGCCAGGCA	ACCGCAGCAG	CCACCAATCA	420
TTGGGATTCG	CGCACTGACC	GACCAACCGT	GTCTGCGACA	CCCCAGCCGA	ATTGGTGGTC	480
TTCCGCGGGG	CGGCAACCGG	AATCAGCCCG	ACGCGCTCCG	CGAASCACCC	TCATAGCCCT	540
ACATAGCAAC	CGGNTGCTCC	CGCAGTTTTC	GGGTTTMTCC	CGCTGCGCAA	CGGNAAYNCC	600
TCGAATTCAG	ACGNAACAAA	TTGCGCATY	ARNGTCTTCC	CGAAGAACCC	ATTCGCTCTA	660
TTCCGCGGGG	GGGTCGCGCT	NMNAACCGG	CGGWAAGCC	CGGCGGCGCC	CGGTTTWTTC	720
CGCTTTCTCC	CGGCGCGCGG	TTTGGTCMCM	CGGCMNTTWN	CGGNTGCGCC	CGGCHAAAAA	780
AAAAAYCKCG	TCGAATYAAA	CGGCTYMAAA	AKCTCGGSSC	CGGMAACCCG	CGGKAACKWAA	840
ANTTAAGCCN	CAAAAAAAWW	NCANNMCCCG	NGGCGCCTAA	GGKYTTAGCG	GTGCTTNANG	900
AAAAAATMTC	TANATMNSSK	TTNMAAAAAA	AGCGGWAACG	CGGNNNKKCN	CGGAWFARPP	960
CGCTTCGCGG	TCGNSLWJLI	KKKKCTTCCG	CGGNNNKKCN	CGGNNNKKCN	CGGNNNKKCN	1020
CGGNNNKKCN	CGGNNNKKCN	CGGNNNKKCN	CGGNNNKKCN	CGGNNNKKCN	CGGNNNKKCN	1036

(2) INFORMATION FOR SEQ ID NO:306:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(B) MOLECULE TYPE: Genomic DNA

```

GCCATGGGCA ACGGCTACTG GGGCAACCCG AATCCATTGG GCGTCTCACC GCAACCCCGG 120
AAACCGGGCA CGCGGGCATG GATCAACCCG CCCACCCGAG ATCCGAAATA GCCTCCACAT 180
AATGAGACAC TGGCGCAAAG AGCTTGACAG GCGCCGCACC ACGCAAGCTG TTAGACGTGT 240
CGGTCTTGCA AGAAGCGGGT TGGCCACCCA AGATCACGCC GCCCAAGGGC ATCGAGTCAA 300
CGTTGCGGTG GTAACGCGCT AACGTGCGCG CCGCCAAGAA ATGACGGTGC GCATTACCAT 360
GGCCCTGCTG ATCACCCTTG GCCACCTGCG CACCANAACAT ATGANCAGCC TTATGCCGAG 420
TCTCGTGGAC ATCGGCAGCC GCTTCAAAAA CTCCTTGTCG ACAATSGTAT TGCTGANCCG 480
CCGAATTCIT NTRCTTGCAA SAACACTNCA TGTTCNSGGT NAACAACCTT GGTTCGAAAA 540
ACANCCAATA TTGAANTCCC ANTGGGGCAM GAACCNCTTM CGGAAGKTGK TGGGAACGAA 600
TGKTGCCCCA AAATCCCGGG NGGTRAAAWW CCCNSNATGG MSAATTTTSC CTNGAACAAAM 660
AAAAGGTCCA AGKYCAAAGG NGCCCCCCCC SGNAATTTGG TGAACSCAKA WYANRTTCCC 720
WWWTCNAAAT MTTNGGGTCC KNNTCCCGWT AAANGGGSCN CCCCNCRRGG GMGTYTCCCC 780
NWNMGGMGN CYCSCCCCA AAAAAAAMMM MTTTCSGKGG SMGGKKCCCC CCSGGTYWGG 840
GKKYTTAAAC CCGGKGGGTN CAAAAAANAN ACCCCCCAMS NGGGGGGAAA ATTTGNAAWT 900
AAGGKKKTKC SCMACCCCAA AAANMMNNCN AWCNCCGMGK SARGGGGRNY TTMKAGGGMG 960
GNYCCCCCW YCGGGGGGNA NAAYAAAAGK NGSNGRGAAT NTTTTTTTGT RSSSRNKTIT 1020
TYNTCTTCN CCMGNRWWG SRAMNTGKTS NSSGGGSGCC 1080

```

2. INFORMATION FOR SEQ ID NO:307:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1040 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:307:

```

AATTCGGCAC GAGCTTCACC AAAGAGCTGA CATGCCGGGT GATGGGACAT CCCATCGAGG 60
TCAATACGGG CATGATGAGN CCCAANGGAN TGTGGCGTTT GCTCAACTGG ATTACGGTTG 120
TCAAGGTGAA ACCTTTTGGG CCCAAAGATG CCACGCTTAA GTTCCGCTTG CACCGTGCAA 180
TGTGTGATG GATGCTGGAA CCGCGCTGAC TGATAANGAA TTGCTGTGTC GCGGGGCACT 240
ATCGATGCTG CAGTTTTCNC TCGCGCTTAA AATTGCTGTG GCATCATCTG GCAGGCTATG 300
TTCCCGCTAC CAGTCAGCCC ATCATGATG TCGCGCTAAC GAANAAGTTA TGACATGGCC 360
AAAGCGAMTC GGGCATSCNC CCGGCAMTTT CCAACCTGCT TGTGTNTGAA CCGTMTCAAC 420
CGAATCGGGC GCTYAAAAGC NGGCTTGCTT TGATTMMAAC CNAACCCNTN CNATYCTTTG 480
CCGNGMNTG CATTCTCTCT AACTCCCKKC SYTCCCTGCG TGAAAGGEMA CTNCCCCCCC 540
TTTGAGCTTA MPTTTCAAA AAMCCGNTNA AGGCGAATNN GAACTTNGCP TCAANTAMM 600
AAANTCGGSC TTYGGNPGC CCGGCAAYW TCGGCTNGGG GNNNTYCTCN GCTTYNGGCG 660
AAACNTTTT CORTNMMN: TTTACAMCCG NCMNTMTTAM TCGGCTNNAG GWCCTGGGKH 720
TTTTCNCAW TONNSKTTT TTTGGGGGCG CCGCTGCTMC NCTGCGGCGG GCGGCKKMAA 780
AAAAACMSA EPCCNCTGGG KTCCTGCGCM RNATNGGGCG TCTPAAACAA ACCCCANRA 840
TTCNMGGGC SMACCGNNGN NNAAAAGCT TONSTMANM MKGMANNCT SGMSCCMNSN 900
NCTGCGGCT TTKGNGAPN AANAMKMGCM CCGNCCGNN BAAAGGGGMS GCGKSCNNGN 960
NGASNWMGN DRNGANRCC NCTGNTNMPN NNGNNGNNN CGGPKNNACN NMKMCWWSMC 1020
NSNMGNNGN CGYMTNKCCG 1040

```

3. INFORMATION FOR SEQ ID NO:308:

(1) SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

AATTCGGCAC	GAGACAANGG	CGTGAATGG	GATCCGGCCG	AGCTGGGGCC	CGTCGTCAGC	60
GACCTGTGG	CCAAGTCGCG	GCCGCCGGTT	CCGGTCTATG	GGGCCTAGTT	ATCTGCGCCG	120
AGCGTGAAC	CAGGGCGAGA	TTTCGGCCGT	TTTCTCGCCC	TGGCTTCACG	TTCCGGCGAAG	180
TKGGGAACGG	TCAGGGTTCC	CAAACCACGA	TCGGGATCGT	GCGGTCGGTC	CAGGACTGGT	240
ANTCCTGATA	CTTKGGTACA	TCGTGACCAA	CTGTGGNCAA	TATTCGGCCG	GCTCCTCGTC	300
NGTCGCGTCC	CGCGCGGTAA	GGTCCANCAC	TTCCCTTTTC	TCGTGCCG		348

(2) INFORMATION FOR SEQ ID NO:309:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

AATTCGGCAC	GAGACACCGG	GTGTTGACC	AACGGACGCT	TGGGCGCGGG	CCGCTTGCGT	60
GGCATCAGCC	CTTCTCCTTC	TTAGCGCGCT	AACGGCTGCC	TGCCTGTTTG	CGGTTCTTGA	120
GACCTGCGCT	ATCCAGCGAA	CCGCGGATGA	TGTTGTAGCG	CACACCGAGG	AGGTCTCTCA	180
CCCGGCGCGG	GCGCACCGAG	ATCATCGAGT	GCTCCTGCAG	GTGTTGGGCG	TCGCGCGGAA	240
TGTACCGCGT	GACCTCGAAC	TGACTCGTCA	CTTCACCGCG	GCAACCTTCC	GAAGCGCGCA	300
CTTCGCGCTC	TTGGGAGTCC	TGGTCTGTCG	CG			332

(2) INFORMATION FOR SEQ ID NO:310:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

AATTCGGCAC	GAATGCGTCC	AGACCGATTG	AATGCTGCGG	CGAGCACCTC	CGCACTGCAC	60
AACCTGCAGC	AAAATGTTCG	GAATGTGGTG	AACGAGCGCT	TGCAGACGCT	CACCGGCGCG	120
CGGCTGATCG	GCAACCGGCG	GAACGGGACT	GTGGAAGGCG	GGGCTGACCG	CGGCGCGCGC	180
CGTCTCTGTT	CGGCAACCGG	CGGCAACCGG	CGTGGGGGCT	GAACGGGAAC	AACGGCGGGG	240
ACGTGGGGAC	GCGCGCGGCG	GAATTTCTTC	GCACGGGGCT	ACGCGGCGCG	CGGCGCGCGT	300
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	360
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	420
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	480
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	540
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	600
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	660
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	720
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	780
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	840
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	900
CGTACAACTG	CACCGCGCGG	GAACGCGCGG	CGTGGGGGCT	CGTGGGGGCT	CGTGGGGGCT	960

NNTNCYTTKN NATTKGGNNA AAAANCCCTY CCWCSGRAC	NCCCCCNGM GRGMCNNTNN	720
NTTTYGNCNN CCGGSNAAM RNTTKATTTG NGGGGNTCN	GGGTMNNNA AACCCCAAAM	780
MNRNNKCSA ANGGGKSGC NKNMMNSGT TTTYCKMRA	MRNWTYKNKN NTCNGARSRN	840
NAAMCNSNK NGKKGKNAKAA ARNNTTWKTN KNSCNNCNN	GRRNGVRGGC CKMKGSNMNG	900
MCWHNAWRNG NNGSNCNCKC NNMNAAAAA AASGGVNCKS	NSMKNNKKKG NRGGGGGGGG	960
GG		962

(2) INFORMATION FOR SEQ ID NO 311:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

X1 SEQUENCE DESCRIPTION: SEQ ID NO:311:

AATTCGGCAC	RAGAAGACGC	CCGAANGTTT	GGCTGGCTC	TACAACTTCA	TCAAGCCCA	50
GGGGGAACGC	AACTTCGGCA	AGATCTACGT	TGGCTTCCCC	GAAGGGGTCT	CGATGCCCCA	100
GTACCTGGGG	GCACCTGCAU	CCGACCTGAC	CCAGGATCCG	GCTTCCAAAC	GGCTTGGCTT	150
GCAGAAAGATG	TGGTTGAGG	TGGCTTGGAC	GATTTTGCAN	GGACGGCCNG	TGACCGGGAC	200
GGGTTTGGTG	TCCGCACTG	TGCTGACCA	CCGCGGCACC	GGTTTGACCT	CGACCACTG	250
CACCACTGCT	GGCGCTGCTG	CCG				323

(2) INFORMATION FOR SEQ ID NO:312:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1034 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

X1 SEQUENCE DESCRIPTION: SEQ ID NO:312:

AATTCGGCACT	GTGTGTGGCG	GGCTTCAGAA	GAAGATGATC	GGGAACATCG	CCAGCGCCCG	50
CTAGGCTATG	GTGCGGCTTA	TGGCGGAGCA	ATGATCATG	AGCGGCATAC	AGCGGGCGCT	100
CTCAGCCGAC	AGCAGCTTGT	GTGACTTGG	TTGTTTGAGC	AAAGCGGTGT	AGACRAACAT	150
CTAAAACGCG	AGGCTGACCA	GGGCACTA	TTGCGGAGG	AGGTTGCTGG	GGCACCATAG	200
CTAGAAGAAC	AGATGACCG	CTNACTCA	CTGATGCA	AGCGCTTTGG	GGTGGGACAG	250
CTTGGGCGG	CGAAGGGGCG	GGCGGGGCTT	CTTTGATCA	CTTTGTGAT	ATCGGCGCTG	300
CTNACCACTT	AGCGCTGTTG	GGCGGGGCTG	CTTGCATGCT	CGCGCGGACN	AGCTGTGTA	350
CTATGAGCAJ	CGATCAATG	CGCGGGGCTG	TGCTGGGCTT	CGTGGCGAAT	TCAACTCGCT	400
CTAAACTTGG	CGGCGGCACT	CGAAGCGGCG	TGAATGATG	AACTTAAACG	GCTGAACANT	450
AACTACATAA	CGCTTGGGCG	CTCTTAAGCG	CTTGTGAAG	GGTTTTTTGG	TTAAAGGAAG	500
AACTATTTCG	GGATACCTGG	CTTNTWTA	GAAAAGGCGG	AGCGCATNCG	CGTCCACAGT	550
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	600
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	650
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	700
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	750
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	800
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	850
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	900
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	950
CTGCTGCTGA	ATGGAATCG	MTNCGGCTG	CTGCTGCTG	AACTGGGCGG	CGTTTTTGTG	1000

SCNSNGGKBC CCCC

1034

(12) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

```

AATTCGGCAC GAGCCACAT CCGGGGCGGC TCGTTGCATG ACTCGTTTGT CATCGTCGAC    50
RAGGCACAGT CGCTGGAGCG CAATGTGTTG CTGACCGTGC TGTCCCGGTT GGGGACCGGT    100
TCCCGGGTGG TGTGACCCA CGACATCGCC CAGCGCGACA ACCTGCGGGT CCGCCGCCAC    150
TACGGGTTCG CCGCCTGATC GAGAACCTCA AAGTTCATCC GTTGTTCGCC CACATCACTT    200
TCCGCGCCAG TGAGCGCTCG CCGATCGCCG CGCTGGTCAC GAGATGCTCG ANGAGATCAC    250
CGGGCGCCCG TGAGTGGCGC TCCCGCGAGC A

```

331

(12) INFORMATION FOR SEQ ID NO:314:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1026 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

```

AATTCGGCAC GAGATCGTCA CCGTGGCGAC CAGTGCACCC AGGCCACGCC ACCAGTTAGC    50
CGTGATGGGC CAGAAGATGG ACCAGGTGCT GGCATCGCCG TCCACCCGAC TGGAGCTGAG    100
TACCGGGATC GCGGTCTCTA GTTACGGCGA TRAGCTGGTG TTCCGCCATCA CCGCTGACTA    150
TGACCCCGCG TCCGAAATGC AGCAGCTGGT CAACGCTATC GAACTGGGTG TGGCCCGTCT    200
CGTGGCCCTC ANCGACAATT CGTGTCTGCT GTTACCAAGG ATCGCCGTAA CGCTTCATCC    250
CGCGCACTCC CCANCGCCCG CGGGCGGGGG CGGCGCTCTG TCCCGACCCG CCGAGCGCGT    300
TACTGACGGC ATCTCGCTCG CGTTTAAGCT CTGAGAAGG TCGCTCGTGC CCACTTGGG    350
TTCGCTCACC ATCNATCCCG CGCGCATCA CGGCTGCTG TTCCACAGA CTTGAGACAC    400
TCCCAAGCAA CTGCTCCGCG AMTNCAGGAA NTCTCTGCG GCACCGCTT TTGCGCTCT    450
CGCTTAAACT TCGNATSTTN CGCGGGGCTT TCGGCTTTC TCGCGGCGCG NTCTTNCBAA    500
ATCGGSMMAA ATCGCCANMC AAACCGCCCG GGTCTTSSUG CGCGCGNGGC CGCCNAWNCT    550
TAACCCCGCC NTAAANTCT TTGKTCGNNN CCGCGGNNNN NCAANSCAN CGCTTTKGGC    600
AATTCGCGCC CCAWTTTAA CGGAKCGGNN AATCGCAAGY TMMGKCCYCY KNAAAAAAAA    650
AATTTGSGSC CCGCAANTAA ATTCCGNGGC CGTTTGGGG CGPANCNYNT TTTMCCSNGS    700
PKGNHNAAMC CGGANCSSCG KAAYTMMTKG NAAVTGCGSN AAMBNTTTTC TAAANCGCCN    750
TNGCGSAAA ATTNNAMAAM CMNKTKGSHG GGGGTTGSHG SKKGRAGGM AAAAAANRBN    800
PTTNMNNNN SANMNCNNNN CGGNMNNNN NNNNNNNNNN TNAANMCGG TCGCGGCGCG    850

```

1026

- (A) LENGTH: 324 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

AATTCGGCAC	GAGAAGACGC	CCGARNGTST	GCCTGGCTC	TACAACTTCA	TCAARGCGCA	60
NGGGGAACGC	AACCTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	120
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCG	GCCGCGAAAC	GGCTTGCCTT	180
GCAGAAGATG	TCGTTTCGAG	TGGCTGGAN	GATTTTGCAN	GCGACGCCNG	TNACCGCGAC	240
GGTTTGTGTG	TCCGCACTGC	TGCTCACCAC	CCGCGGCACC	GCCTTGACGC	TCCACCAGCT	300
GCACCACTCG	TGCGCTCGT	GGCG				324

(ii) INFORMATION FOR SEQ ID NO:316:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

AATTCGGCAC	GANGCSTGCC	GTNAACACC	AGCGCGGGC	TGCCAGATAT	CCCGGACTCG	60
CTAGTGGCGC	CGGTGGCGTC	CTTGGCTTCC	TGACGGGGCG	CGGCGACCAT	AAGGTGCGTM	120
ATGCGCGAGT	AGCGGGCGAG	CTGATGGAG	TCCATGATGA	TGCGACTCTC	CAGCTCGCGC	180
ACCGGGAGCT	TGGCATCGCG	CTGATCAGC	CAGGACCGCT	AGGACAAGTC	GATCGAATGC	240
ATACTGGGCT	CCAGAGTGGC	CTTGCAMTTC	TNGGCTGCTC	CACGGCAAAT	GCCTTGATTT	300
CTACTGGGCG	TANTGTTGCC	GCATCGGCTC	CGGATGGAAT	GGGAACCGCA	GGATGGCGAC	360
GAACGGGTCT	GANTCGAGCT	CTGGCGCTTT	CGGACAGTC	GTGACACACC	GGTACTCGGC	420
ATAATCTGCG	CGGNAATGCG	CGTGGGACCG	TGCGGACNAT	AANAACGGGC	ACNACAATCG	480
CGCGCGCGCT	CAGCGNAACA	ACANCTTGGC	ATGGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CGGGAACGCG	TGTTGGCGCG	NACTTTTCTT	NNAWTAACTG	CGGCTTCCGH	CGCTGGNGCA	600
ATAAATGGGA	AACCGTTNCC	CCAGCTTGAA	GGGCTTCTTC	NATTTTACT	GCTAACCGCG	660
AATTTTGGCG	GANTCGCTCG	CGCGGGTTTC	CTTNTTGGCG	ACCTTNGNAN	GGCGCGGCTA	720
AGTTTCTGCT	CTTGAAGCGG	GAAGCGGAA	CTTNTTCTTC	AACCGCMNAA	MYMTTTCGCG	780
GAAGTCTGCT	CGCTTTTAA	GAAGCGGCTT	GAAGTCTGCT	GGGTAAGAAA	GGGCGNKTTC	840
GAAGTCTGCT	GGGCGGAAAG	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	900
CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	960
CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	CTTCTTCTTC	1020

(ii) INFORMATION FOR SEQ ID NO:317:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:317:

AATTCGGCAC	GANGCGTGCC	GCNAACACC	AGCCCGCGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGGCGC	CGGTGGCGTC	GTGCTCTCC	TGACGGGGCG	CGGCGACCAT	AAGGTCCCTM	120
ATGCCCAGGT	AGCGGCCAG	GTGCATGGAG	TCGATGATGA	TGCGACTCTC	CAGTCCSCCG	180
ACCGGGAGCT	TGGCATCGGC	CCTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGCCT	CCAGAGTGCC	CGTGCAATTG	CNGCGTGCTC	CACGGCAAAT	GCCTTGAATT	300
CTACTCCGCG	TANTGTTCCC	GCATCGCCTG	CGGGATGAAT	GGGAACCGCA	SGATGGCGAC	360
GAACGGGTCT	GANCTCAGGT	TTGCCGCTTT	GCGCACAGTG	GTGNACANCC	GGTACTCGGC	420
ATANATCTGG	CCCNAAATCG	JCGCCGACGG	CGCCACNAT	AANAACGGGC	ACNACAATCG	480
CCGCCCCGGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC	TCNTCCGGCG	NACTTTTCTT	NNAWTAACTG	CCGCTTCCGK	CCCTGGNGCA	600
WTAAATGGGA	AACCTTINCC	CCACCTTGAA	GGGGTTGTTG	NATTTTTACT	GSTAACCCCG	660
AATTNTTCCG	GANTCGGTGN	KCCGGGTTTT	YSTNTTCCCC	ACCTTNGNAN	GGGCCGGGCA	720
AGSTTTTCTT	SYTGAAGGGG	GAAACCCAAC	TTTNTYTATN	AACCCSMNAA	MYMTTTYCSG	780
MNAASCNK	CCCTTTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCSCCYMANG	GGGAGGAAAA	TCTKTCNNCC	GGGCCKAAAW	ACCCMMMYGN	GTGKKKXKSS	900
CCSAAATTTT	NMMRAACTGN	GGGGCCSSGA	NNTTTNAAAG	MSCCCCCSNN	ESTGFCCCN	960
NTTTCNNAA	NMKKGHNWNM	BNMNSCSNGG	CKYNSGGGNN	NNAAGMGGGG		1010

(2) INFORMATION FOR SEQ ID NO:318:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:318:

VENGGGGWNS	NTCAVCAYCA	VCACSGGGGW	WATTCCGGG	CGCAWCTTGT	MAASAGATGT	60
JGAAYTCGGC	AMGAGGGAMT	CKCTMGCCNC	CTGTGGCAAN	CGAATRAGGC	CTRATAATTV	120
CCACTCCACA	AAAAACCTTT	CTGTGTATYT	ACCGRAAATR	AAGGCCSCCG	TNTCAACWYC	180
CCCGGTKTTY	CCRATYCCCG	TKTTGTAMCT	GCCKGGGTSR	AAAYCCCCCG	TGTTGGAYCC	240
CCGGATTGAA	ACTGCCCGKT	TCAAACTGCC	TKTTTSGGSA	CCCGGKWATT	GAMSTCRCCG	300
ATTAAAAAAC	CCGKTTTGGN	CTGSGNCCTG	CGAAATNCGR	AYCCRATAYC	CCATGCCCTG	360
ACTCTCTCCK	CCGTACCCCA	RAYCTGGGTA	TCCTATACTG	JYCCCTAAAK	CAAAWYCKGG	420
CTCTYIMMTH	TTGCKGGGCT	JNAAATTTAG	CAAAATCTCT	TCCTTCCATA	JNAAACNCG	480
CTTCTTCTCT	ACMCCGAAAA	AAAKAATAAT	PAKAAGGTTT	JATNYGAAAA	ACGNCGGGCTN	540
CTONANTNEN	ATCCONTNCC	ACGNCGCCCA	GGGTNAAGTH	TYSGGAATTT	CTMMAACCCG	600
CAANCCGCA	CAACNTNCCR	CAABAAAACC	CTYCNCGGGG	GYCNWNCAAA	ACASONTTAT	660
CTCTCTCTTT	CTGGMWCCCT	TCCTGCNAAA	ACCGAAABTA	TTTTYTGGGT	CCNAGAKAAA	720
CTNCCGGGCT	CTMCCGCAAA	CTATATCTCT	CCGCAANCCG	CGAAACCTTT	CTMNAACNCK	780
CTPMTCTCTT	CCGYSUCAAT	CTGYCGGRAT	CCGSGCCCTY	TCAAAKKKKS	CAKWWNNGNG	840
CTPNNACGMA	ACCGGAAGTY	CCMNAAAAATN	CKCCCGGCTC	CNAACACGK	CTVTCGSAAA	900
ACCTCTNCCG	CCCGGCGRAA	ACCGGCGGNA	PKANTNCCCA	AAAACTNYNG	CTCTCCGCTG	960
CAACMAAAA	AMCCGCGGSM	PMACCTGCTN	MMGCGGKYY	CTTTTTCTTT	CTCTCTCTCT	1020
CACTCAMWGY	CKKINMAAAA	CTAACATANT	CTTCAAAATY	CTCTNYWPSW	CTCTCTCTCT	1080
CACTCTCTCT						1140

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1251 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:319:

```

GGGGGGGNNN NATACATCWT CYGTGYACCG GGGMTCTAKT GGGGGGGGGG AATCTNGTCA      60
ASAGATCTCT NAMTTCGGGC ACAAAAACWT GACAAASYMT CGNGCNMTCC GTGTCTTNKA      120
TCGCAAAACG NGTRACASAC ASACACRTAT GTGTGCCCCAC CASCAYTCK TTGGGACCTC      180
GCTRACCGGY TGGGGRNACG CCACGYTCCG CWTCTATCCC RACGCGGGCC ACGGGYGGGG      240
ATATTCCAGG CACCCACGCGG AGTTTGGTGG ACAATGCCCT GGCATTTTCC TCRAANTTCC      300
TGAAACCGAA TTGNSMTTGA ACCNCCAARG CCCCNCNCR AACARTTGGG WTCCGCGGTT      360
CTGCGGACCG KTTTCCGGGG GTNTCGGCAN AANCGCACCC WTGGWTTCTM TCNCCGCAAC      420
GGGGGGACAA NTGCGGTTGG AATTTTGCRA AYCGGGGCGG GGATTCCSCA AACGGGTGCG      480
GAAACTGTTT YCRAAMACCG GGAACCGCAA TTTCCGGGCR ANAAATTTCTN YCNACCACT      540
GCTTCTACTT CCGCGACCGT AACMANTTTC ATCCTCTNNN CCTCTGCCCT TGGGGCAGGG      600
CKAAAYACCG CMTTKGGTIT CGCAACCTCC GGGCCAANTC CCHAMCCRCA CTTCNATTT      660
GNTGCAATT GCGCGCGGT RANAACCGCC NTGGCCNNTT CCGASSAAAA NCGGCGCTNT      720
KGGCNSCCCC AGTAANACCG TACCNNAVTS CAWCTCTTGC CAAASTTKGG ACGAANSKTG      780
GGNTTCCGGK ATTTTYTTGS GGNCCGCTN TATNGGNTN GGGCKCYNC NCSTKTGKCA      840
NASSKAYCCG NGNKGGGGGT ACCCGGCTMG GGGGTTTTTT NSSGCCCCCC AWAYGNKSTG      900
GCCCCGNNCG GGAAKAATWT MWTMCNSGG GGGAAWTTTT NTSTGGAMCS SGGACYCCCP      960
GGGGGKTTTT TCCCCNCSSA MNAWANGGGG GGGGGANAYT NTGNSGNGGG KWNTTTTATT      1020
TUYCYCTM TKACMSGGGG GTTTKKAENG GGGGGAGAAA ANAAAAAAA RAKGGYKNTT      1080
TGKACACNTT GKNWNWNR NAGAGTCTT CKCKCCNCGG SNTTCTTT MGNSGSGYGG      1140
DRIGNNNAAA ACNKRMMAC KCSYTYCCG CSYCTCCTCC NCNCGGGYGS NGSCONSTIN      1200
NNKCRHNTA TMTGNCSTN SCCTCCNCGG GKNKWTCTG TMTCNMYGGG      1251

```

2. INFORMATION FOR SEQ ID NO:320:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1099 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

11. SEQUENCE DESCRIPTION: SEQ ID NO:320:

```

TATTTGAAA GAGATATCAG TAATCTTCTT AGGCTAATAA AATGGAGGTA TTAATACTTT      60
TATTTGAAA TATACATCTT TTCCCGCTAG KTGCTGCTTT GATGCTTGGC CTGCGCTGAK      120
TAAAGCTTCT TGGGGRNACG CCGATCGCCC CAAAGCTTTT TGGGAGGACG ATCAKCGCGG      180
AAATATTTAA TTGNCGCGCG ACCCGGGGCT GNCATAACT TAAAGCTGCT GGCMTGCTG      240
CTGCTGCTAT CTTCTTCTTA ANTGGAACTC AGCGGCGNMA TAAAGCTGCT TAAAGCTGCT      300
TTAAAGCTTA TTGAAAAAAT CTCAAGTWTG GCGGCTACTT TAAAGCTGCT TAAAGCTGCT      360
TAAAGCTGCT CTTCTTCTTA TCAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      420
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      480
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      540
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      600
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      660
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      720
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      780
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      840
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      900
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      960
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      1020
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      1080
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      1140
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      1200
TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT TAAAGCTGCT      1251

```

YVNGSGNKMY	ATCATGTTT	YVNGSGNKG	MTGATTGCG	AGGTAATTT	TGTMNASAGA	1
TGTGGAAYTG	GGCAGGATTA	TGTGCGGCG	TAACTGCGAA	AGTGTGHTAA	TGTGATGCG	100
YVNGSGNYCA	AGGAGGTGTT	YVNGATGCGG	GCGAMWTVCA	AAGGATTTAT	TGTGCGGAG	100
AAAATTTTMOG	GTGTTTASCA	AGTGTGAGCG	AGTGAASCAA	TAGGCTTTTA	AGTGTAAATY	100
YVNGGTTNKY	YVNGGCAACA	AGTGTGATTA	TGCGGCGCGG	AGGACAAAAA	TGTGANTNGT	100
YVNGGTCRAA	YVNGGTTTCT	YVNGGCGTAA	ACTGCGGTAT	GCTGTTGWTG	AGTGTACAT	100
TAAACCGCGG	GCGGTTGTTG	YVNGGCGGAA	ATGCGGTAT	TGTGTTGWTG	AGTGTACAT	100
AGTGTATGTT	RAAGCATG	YVNGGCGGTA	AGTGTGCGG	YVNGGCGGTA	AGTGTACAT	100
YVNGGATGCG	GAATTTACGG	TAGGTTTAA	AGTGTGCGG	YVNGGCGGTA	AGTGTACAT	100
ATGMMWNGCG	TACGTTTAA	AGTGTGTTT	AGTGTGCGG	YVNGGCGGTA	AGTGTACAT	100
TGCGMGGGG	YVNGGAAKTT	YVNGGTTT	YVNGGCGG	YVNGGCGGTA	AGTGTACAT	100
YVNGGCAAG	YVNGGGAAT	YVNGGTTT	YVNGGCGG	YVNGGCGGTA	AGTGTACAT	100
YVNGGCAAG	YVNGGGAAT	YVNGGTTT	YVNGGCGG	YVNGGCGGTA	AGTGTACAT	100

12 INFORMATION FOR SEC ID NO:323:

(A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

x1: SEQUENCE DESCRIPTION: SEC ID NO: 323:

INFORMATION FOR SEC 10 NO. 121

A LENGTH 1210 case path
B TYPE 0000000000
C TRANSFER 000000
D 000000000000

A. SEQUENCE DESCRIPTION: 000 10 100 004

[illegible]

TCCCKTTCGG	GGGCGCGCCN	AAAAACCACC	AATYCCGYTG	GGGGTGKYCC	CMCAGGCSGT	480
TGCTYCGNGY	CACCTGGCCA	AAVYCCCAWT	AKATTGGGTG	SCYCKTSCGG	TTSYTGGGCY	540
CAATTACCCC	CNCGGNAAA	GRRAAANA	ATCNTCCNTT	TGCTCGGYCA	YCITTTMTTG	600
SAAAAGGGGC	ATGGCSCGGT	TYTTTTACCT	CAAYCCCCNA	NCANTWACCT	YTCCSCCGG	660
GGGNCANAA	CGSTTNGCTC	CGSGGNAKCC	TKGTMCCCGN	ATCNAAAGGC	CNGAATTTGG	720
TYVSSTYCNA	ATTWTWKKY	CCCCWCNTTG	YAAAAAKCCA	AAASAKCCCK	YCNCAMMYKT	780
NGGGGTYSSG	GCKNYCTTK	SNMTTAAACC	CYCCCCAAAA	YVNSGGGKKT	TCCGCYNSAT	840
KCCACCNCCK	GNGGGGGGNA	SAAAAAAAY	TTTTCCSAAA	ATCCCACCYY	TCYKTKSTRY	900
AMACCCCTT	TYMKKAYTC	CKYGCNATTC	SGMTTCWAAA	TYCCGYOGCT	TNTTCCCCCK	960
CSGGNGCCCC	AAWTTTGKTT	YNCNANTTYC	CCCNAAAMNCM	AWTMGGGGKS	KCCATTCTGG	1020
SCYTMAANTA	AAANAANGGG	NKTTTYCTY	MANAAACACN	GTGKCNCCN	CNAAMAAASN	1080
AKMAAAKAGN	KHKMTKNNSA	AANCCNCCCC	CTSTYTNYYT	NKTNMNCCKC	CYGGKGNKGM	1140
SWSWYNTTCT	NCCCRCCCC	YNYNKTGANA	AAMMNCYCCS	GGSTMCRNAN	ASNMTTTTCK	1200
STSTNMGCC	KMBASNANAN	MCAMWKWYCC				1230

2) INFORMATION FOR SEQ ID NO:325:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 1022 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

1X) SEQUENCE DESCRIPTION: SEQ ID NO:325:

NGNBSGKDMA	TMAKWTCTC	ACSSGGTCTA	TGCGGCGCAW	CTMGTMAASA	GATCTCNAAY	60
TGCGCAMNAN	GCATMTMMC	CATATATAAC	CATTGCGTCS	GYWTSCAWCT	CRAAWCTGTC	120
TTTESKGGGG	TTKTACRAAG	GTGGMWTGYT	CWTYCTTRAA	SCCCTCRATC	TCKTKTATYC	180
BTKGGGGTYC	ACTTTAACSG	RATKSCCTGG	TTKTAYCATT	RATGCAAWTA	WTGGYCRAWT	240
TTTSCAGGGC	RACGGGNYCT	TTTYCCGCPA	BRACAATNGA	TTGGAWYCGJ	TYCGCRAGGC	300
CGGBCACCAR	ACGGGGCNC	AAAGGYCCGC	GCAAWTSCCT	JGKTCAAAAA	TGGTGCRAAC	360
GAAMCNATCG	CGGVTTRAC	CGCAGYTAMC	ACAAKAAAAAT	TCCGVTGGCT	GCACCAWNNI	420
TTVIRATOWY	CGYCCCGACG	TTTAAATTTGK	VTGCGGTATT	GCCTKCGCTG	CTCRACAGCM	480
TCNCCCKTCA	AACTGCGGT	GACTCCAACT	GGTCTGGYCS	AASGGGGGYT	CAMCGGACAA	540
AACCCCRANW	TCGCCAAATT	TTCCGCCGCC	TCGCGGAAAN	GKTGATMTT	TCGNAACCSA	600
CMGGGNNYTW	NAACCCGTGA	CGSSGSKNKA	MYNSCCSGGA	ANTTTTCCCT	TYNGGGCGRN	660
AAANCCCTTT	AAGGTACCC	KGGNGGGGKG	GGGYTTGGG	AAACCAACCC	CKATTGGKTT	720
TGGAATNTT	TKNCCGCCA	TTGNSGGGGG	GGGCGGCAMT	TCMCGTTTTN	TCMCGMNTVY	780
CTVYGGGAAT	TNYTCCCGCG	SAAYYCGGSM	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	840
TCRATMAWWT	TGTTTTTTC	CGGCGNCCG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	900
TCMCGMNTVY	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	960
TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	1020
TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	1080
TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	1140
TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	TCGCGMNTGG	TCYVCTTAA	1200

1) INFORMATION FOR SEQ ID NO:326:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 1033 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:326

```

NNCGNNKNTA TAMAYCWYCT NCACCSGGGA TCWATTGCGG CCGCAATCTT STMAASAGAT    60
CTCKAAYTCG GCAMGANCCG CAWCTATTTG KGTGRASCGC ACCAGCGRGA CCTCGCSGKT    120
CKTTYCTTGC AGRGAGGCGK TGGGTGGCRC CGGTGGCAAT GCCAACCGCC CCCCCAACN    180
CCGCAAAATMY CRAAAAACAA CCCSGGGSTA GKTCCSGGCC GCCAAATMAA TAACCGTKIT    240
AACKCAGGCN ACGGCCAACG GGYCCCGCCC AACCAAGCNA CCTCCCCSCC NATAGGYCCG    300
GTGGGGGCTG CCKTATYKCC AASTCGTCAY CTCNACGGGM CGGYCCMCWT TCCGCTCAT    360
CGGTCTCTCC TTMATTTTTC CRTCCACYKG GCGGGGAACY TTTTNYCNC CTTGSCMAN    420
CACCNAAAGGY CNAAAATTNC CMTGCCFYG SNNCAAAYGR GATTGGGGTY CGKXTTTTNT    480
TCNMCCMAAC CCCCNTTTNA CGCCCCMATC CCYTWATACC CCCWWMCMNS ANGKTTGNSA    540
AAKTNNCCCC AAATRCCAAA MTCTTCGCC NTTMTWMCY YYCCTTTCCC CMCCCNAAA    600
GGSCCRCCYY TCGGGAANTY TCCCNCAAA AWTCAWCCM TTTCCCNCCA AGAAWTTCSG    660
SACTCCTTTN TTCNGGGNAM ATANATYYTT YCKTNGGSK TCCGMTNC AMMAATNTCC    720
RGGGKAAMCC AGKNTNNTCC YYYCCCAA NNTYCCYKG RMCYNYCY TTAANRASE    780
SAACCCSGG GKCYNCSS TARCCCCAM KAAATTTCC CCCCSTTTC TYNNKKMRW    840
GCCCCSAAM ACTMTWAYTT TCCCKCGNNN TTSYCKCS KCAMWMMTG KKNCTTTTT    900
YCSMATAMA CTINGGKCT NTCNYSGCG CMAAANAAGG CGCGSTCTN TTCWMAMACA    960
YNTSGNMMA SAAKAKWATA AWWNTRKYYK TKNCCCNCC CKKCTTSNN TNKCCMCSKS    1020
GGGKNNKKR CMTCCWNC CKCCCNCAK CCKWATMCCC CCCCCKCCGM NCMNTTTTKT    1080
CCC

```

1083

(2) INFORMATION FOR SEQ ID NO:327:

(I) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1069 BASE PAIRS
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: GENOMIC DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:327

```

GGGGNNKYAT MCAYCWTCT YACSGGGMND TATTGCGGCC GCWYTNCTM CASAGATCTC    60
GAAYTCGGCA MGAATAAGW GATGTGCTG ACCTTMCCE GCGGGACGER ACCRACAAAG    120
RAASCAGGCC ANAATATTGG CCACAKITGG TCACATATTT ACCCAATTMT AYCAGGGAYT    180
MCCATTCCGK GGACCRACCS CACAATCCCR ATSKTGSTTT GCRAACCCTR ACCGTCCCCA    240
MYTYCGCCRA STTGAACCA GGCRAAAAA CGGCCRAAWY CTCGCCCTGA NTCCCGCTCS    300
GCGCNAATAA CTAGGCCCAT TKAACGGAAC CGGNGGCCSC NANTTGGCCA ACAGGTCTTR    360
ACAAAGGGGG CCASYYCGG CCGGTCCCW TTYCACNCC TNKTCTCKTG CGGAATYCGG    420
WTCRATNYC CWTGGGCTT TKTCHYCKYC KYCGGTNCCA AWICTNCGTA TNJATRKG    480
TCCCTAAAT SCANATDGG GKNYCCATT NCTGGSNTTC NATTTAMMAN SRRCGGTCT    540
TTCWITCRA AACCGSNTGS GCGNNMCCA AAAAAATGAT ATAATAATSE YGSCITGAA    600
ACCCCGGCCC CCATTTRWT CSGTTCACG CCCCNGNGGT TARGATGGGA ATFTYTNAM    660
YONARGCCCT NATTTGSONA AAAACCYCY GGYCTEAAA CNYTTTTT GSKSNTCCG    720
GCTCTTSCG CAAACDAAA ATNTNTNYG GGYCKTNAA ACMGGYCRG RCESGAARTT    780
TTTTYGSTTC AACCCCAAC TTTCAACCC NTITTYTYT TCCGSCSMN TNSSSGSNT    840
RSSSNTTTCY RAKKCCNNN GGGGWAYCN CCCCMTTTC TTTTTTTT CCGTNNMAM    900
NGKITCTICA AASMCCDCC SCGCCNSAA ACCCCCTNAE GTTITYCMA AANNWYNGN    960
NNCCCCCCC MGAATAAAY YCSCCCGNR ACSMSNGSA MCCCCGGGN NTTRKTTTTT    1020
TNCMSGYCCC CORMASYTT TKAMAWNRE GAMNSMTTY TNRGNWY

```

1069

(2) INFORMATION FOR SEQ ID NO:328

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1210 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

```

NGNGGGGKWK  MATACATCWT  TCTTCACGSG  GGATCWATTG  CGGGCCGCAW  TCTNGTMCAA      50
SAGATCTCGA  TYTCGGGCAM  NACCCACCWC  TCCRAAAAAA  ACCCRAAWCT  CGGGSKCTYC     100
GARAAGTGTT  GCGCGCKTTR  AATTTAACAA  ATTCAAGTGC  ANAGTGTCAJ  GGCKTTACWT     150
YCCCGGCAAA  GGGGOCACAA  CCTGCAGRGA  SCACYCRATG  GKTGYTGKTB  CNGCGGGGGG     200
CGCGKTNAA  GGACCTGCTT  GGGTKTGCS  TMCAAAANATC  WYCCGGGGGT  YCGCTGGRAT     250
MCNCAGGGGT  CTCAAAAAAC  CGCAAACAG  CACCCANCC  NTTTACGGG  GTTAAAANGA     300
AAAAGGGCTG  ATGCCCCCAA  GGGGGCCCC  NCTCAACCTT  CCCTTCTCA  ACAACCCGCT     350
CTCTCKTGCC  RAATCCGRWT  CCRATNYCNC  CWTGGCCCTT  TCKYCTYCT  CGGTACCCAA     400
ATCTGGGTAT  CCTATASTGT  CCCCTAAWTT  CCAATCTGG  GCTGTCTAT  TSCCTTGGCT     450
TCCAAATTTA  CCANCAAGS  TTTCTTNCAT  NCCAAAAACC  GNTKGSKCC  NRACCCRAAA     500
AAATGAATAA  TAATAANNG  KNNNTTYCNA  ACNCCCCCC  CCGNATTCA  TYSNGTTCA     550
NMNCCGCCAG  NCGKTAGGK  GGGAAANYC  TCMACCYCA  ANCCCTWAS  TTTTNGRAAT     600
KAAACCTCTC  YCNGGGTOW  TYMAAAAAA  NTTATTTGGN  NGNTTTCGG  MWNCKRKNT     650
SCCAAAATCC  MAAATANTTT  YTTGGTYCNA  TWAAAAAMCG  YGNCCMNCC  GGAAAAATTT     700
TCTTGKTTSA  ACCCCAAAAC  YTTTTCMNA  NCCSKTTTT  CYTCCGCC  AMNWTGGGYS     750
GGGNAKGYG  SCYTNTCTTA  TTKTYTYMT  CMGGGGGGNN  MKMTCCMMCC  CCMTTTTYYC     800
NYWRTTTTTN  KCCCCNTNMP  NNRAANNNGN  YTCNNANAA  AAGCNCNCC  SCCNCCCTNA     850
AAAAWCCCN  NNNABAKTNT  TTMKANNNMN  SKCNKNGKY  YCCCCCWC  YNMNNAAAAA     900
AAATMYCCNC  RASANMCASM  NMGGRGNPST  TCCCGCCSTT  NNNNTMTNT  TTTTTCGSA     950
SACCKCCSC  MNNAAMKNCY  CTTTTYCN  NNGNNGNWN  NGNGMNCCK  CENAGAAMWK     1000
CTKCTCCCKS

```

2. INFORMATION FOR SEQ ID NO:329:

A. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 1105 base pairs
 B) TYPE: nucleic acid
 C) STRANDEDNESS: single
 D) TOPOLOGY: linear

B. MOLECULE TYPE: Genomic DNA

C. SEQUENCE DESCRIPTION: SEQ ID NO:329:

```

CGCCSHNNNA  TMCATCWTCT  GACCGGGMT  CATTCCCTT  CCAACTNCT  MAAGAGATCT     50
TGAATTCGGG  AAKANACACC  ACCCGCTGT  MTATACACCG  CAAATCTTCT  GTKTCGAAA     100
AAGGAGACCG  CCGCGCCGCG  GGGYTCCAAC  GCKTTACYTR  ACCCGCCAGY  TCACTTTTRA     150
AAAGCGGTGT  RAGCGGCGCA  CTCACGTAA  ACCCTTTAK  GAAGFAWTE  SKTGGGCCCC     200
AAGTACCTGY  TCTCWTCT  CTGWCCTCT  CTAGCGCCCG  CTANCCGCG  CTGCGGCTTC     250
AMGACGCGCT  CTCTATCT  CTGWTCTCT  CTGCGGCGCT  CACTGCGCA  CTGCGGCTTC     300
AAAGTAACT  CAACTCTCT  CTGWTCTCT  CAACTCTCT  CAACTCTCT  CAACTCTCT     350

```

ATTTCSGRAA	SAACCCCTNY	CCCGGGTTTT	YCCTGCTCMG	GCSCAANACC	CCCGGGGAATC	660
AAAAASGGTC	GGNCAAANGG	GCMAAAACCCS	SACCCMACTT	WTTCCRCCTN	GGGGGGGSCWN	720
CCKNGTTTTAA	AWKSCCTCYV	CTSCCCAAAY	TGGGKCMAAA	NNGRKTITGGK	TTNGGCNACC	780
NTTTCCGGKC	CCGGGKGKKG	WGKYCTMNMA	GSTTTNTTTT	SCCCCYKAAA	NYSCCCCCCC	840
CGSSSCCCC	CCCGGGGGGA	NNTTTTTAMA	GKKTYYCCCT	CCCCMAAAAA	ANACCCGNYC	900
CCSGGSCCCT	TTKRWAAMTN	KCTSCCCCGN	GNNGGGGKCM	GGKTTATTMT	NNNCCSCCCC	960
TCCGCGSAAA	AAATAKMTT	SYCCCCCNC	CTCCKNCKNR	GKAMSMSCGC	TCCCYCTCNC	1020
GCNKNTWAAN	ARSNCCKKKN	CCNICYCCGS	NSNGKCNWCD	NCCSTSSNCT	NKGCTCKNCN	1080
KAAANAAYNC	NGSMSTSSME	CNKCC				1100

(2) INFORMATION FOR SEQ ID NO:330:

(2) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 936 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:330;

NGSNSNKNNN	TAMAYCWYYC	TSCACSNNGA	ACWANTGCGG	CCRMAWCTNS	TMKASAGATC	60
TMGAAYTCGG	CAAGAGCGCG	AAGAGTGTGT	GCATCTGGTC	ANAGTSTMMA	CRGGGTGCGG	120
CGGCTGKGTR	GACCAACMCAT	NTGCGRACAC	CAAAACCCCTC	CGGGGYCACC	CGCKTGGCCT	180
GCAAAWYCC	CCAGGCGCAC	TCRAACAAYW	KCTVCTGCAA	CGCARGCGCT	TYCGGCGCCG	240
RATCTCTGGK	CASYTCGCG	TGCGGTGCCC	AAGKTACTGG	CSCAYCAAAA	CGGCTCTGCG	300
AACTRAACKT	AAWYTGCGG	AAATTGCTTC	CGGTGCGGCT	TGATAAAATTT	NTNAAGCCAC	360
CGCAAMCCTV	CGGGCKTCTC	CTCTCTGCGA	ATYQURWTCC	RATAYCGCGA	TGGCGCTNKT	420
KYCTYCKYCS	GTACCCAAAT	CTTGGGTATC	CTATANTKYC	CGAAANRCA	AWTCTGCGGK	480
KTCBATKTC	TGSGKTGCPA	ATTTAMMACA	NCGJTTTWT	TCNTACCGAA	AAACGCTGCG	540
CGCCACCPA	AAAAGGATAA	TAATAAKGTG	GNWWCAAAA	CGCGCGCGCG	ERTTCLAYTG	600
GTCCARCACC	CCANGNGGT	AGGTNGGAAT	TCTMAAGCGC	CAJCGGATAA	INTTNSGMA	660
AAAGCCCGCN	GGGYMYCAA	AMMCTTTTTT	TGGMTTGGGG	TCATGKYKCN	AAAAACGAAA	720
TMCTTCCGCT	CGWAAAACG	CGCCCNCGCG	MAAAATTTTT	SKCAACGCTA	AACTTTTAM	780
CGNNNTCTCY	GYCCNSACAA	TGGSGGCKN	NGSCTNTTCT	TWTTTTYNNA	CGGGGGRRC	840
GNCCCGNAAN	YYCCNAANKG	NKCCCGSNMA	AAAGAGANTT	TCMKAAAAAC	CGCCGCGCGG	900
MAAAVACCGC	MAAAKWTTCT	AAASMSGNG	YCGCGC			930

2 INFORMATION FOR REGISTRATION

SEQUENCE NUMBER 10117

- A LENGTH 142 base pairs
B TYPE nucleic acid
C STRANDEDNESS single
D TOPOLOGY linear

MOLECULE TYPE Genomic DNA

[illegible]

Figure 1. The effect of the concentration of the *Agrobacterium* suspension on the transformation efficiency of *Agrobacterium* strains. The *Agrobacterium* strains were grown in YEA medium for 24 h at 28 °C. The cell concentration was adjusted to 10⁸ cells/ml. The cell suspension was mixed with the plant tissue and incubated for 24 h at 28 °C. The plant tissue was then cultured on the selective medium. The transformation efficiency was determined as the number of transformants per 100 mg of plant tissue. The data are the mean ± SD of three independent experiments.

SCGGRAASCG	GTGCCAACCC	RAAACNCKTT	GGGCACYCGG	KTSRACTTTA	AASGGTAATC	300
TCKTCCTCCT	GGGCTATGCT	GGGCCACAAA	CCTSYTGCCG	WGGGTCTGGC	CCTGGGYCAC	360
CGYCRCTTTT	TATNTNTCK	YCTACACNCT	TKGGTYCAAC	CAACCCACTT	CACMAAATTG	420
TTTTGGGKTG	GGGSSGCCGG	YTGTNNCCGK	TAATAATCSG	NTGKTCGGCC	MYCACC GGWA	480
CCATANCCTG	GGCGGCSCTG	GCAAATTTCC	SAAATCATYT	CCTTCTGRAC	CCCCACAMRC	540
CTNSAAATCC	GRATCAATNC	CCCNKGGCTT	NTCYCTCTCN	GTRCCCAATY	TGGTTTTCTAT	600
RKTNCCCYAA	TSCAATTGGS	TTYCCRTTSC	YGSTTCCAAN	TTNACAAMAS	GGTTTTYTCMT	660
ACCAAAACCC	NTGGSCCNNA	CMNAAAAKNA	RAAAANAKGG	KCTTTYAAAC	CCCCCCTAT	720
TCAWYCGGTN	CMRNWCCCCG	NGKAAGGKGN	GAAAYTTHRA	CCCAANCCMT	ARSTTSGNAK	780
AAACCCYCG	GGGTSMCAAA	MKNWTWTTSSC	CTTCGGMCTT	YCCAAATMSA	AAATYYTCKK	840
KRMNAAAAMC	YGNCCCCSAA	ANATTTTTGT	NAAMCCCKMA	YYTRTTWMCC	WTTTTCCYCC	900
CCMCNNSNSG	GNTNCCCTTY	TYATTTCYMM	MCRNNSGACN	CCCCMNTYTT	TWTTCKCWCN	960
MMARGSNNTT	RGRMMNMNCC	CCNCCCCNAK	MTCCNCAAAK	NTTTNAACNN	NNKYCKCCCC	1020
CCCMWMNKNC	CCCCMNCMTT	TM				1042

(2) INFORMATION FOR SEQ ID NO:332:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(E) MOLECULE TYPE: Genomic DNA

(F) SEQUENCE DESCRIPTION: SEQ ID NO:332:

NNSGSGMKKK	ATAMATCWTT	CYSYACCSNG	GMTCWATTGC	GGCCGMATTC	TNGTMAASAG	60
ATCTCGAAYT	CGGCAAAANAK	ACSCMAYGTC	AAGTGTRAVY	CGGTACACATA	TCMTCCGCGNG	120
TCAACMCCAA	AGCCGNGTCA	CCGYCTCCCT	GGGGCCGCAC	CCCCATCGGT	RATGCAACYT	180
GGCCGCGCAC	CGYCAAAAGG	KTCWTTTRAG	CGCTAAAGGT	CAMCAATTCC	TRAGGTVMCN	240
TACCGTTNNT	TGGCCGCCCC	RAWTYCTPAC	CCGGAATWTC	GGTAATCGGR	AATTTGGGCW	300
TCGGCTTGGG	CAATAACKTN	TTGGGCAACG	CGGURWTCYC	NCTGGCCGGA	ATTGCCNCAT	360
TCKTTAAAG	GKTRACCCCT	TTTCCCGGHT	CGCTAAAYTC	CTYCTTGGGC	CCCTTGGGCC	420
CPNAGCASYV	CCCTAACCGV	CMCCAGGCAG	TACCKTTGGC	TTTRAACCAC	CGGRATNAAY	480
TGKTACCCAC	YTCMASSOTS	CTGRANTTRK	TNTGNTGRAA	AANMCCACCN	AACCCGGNTT	540
RATCTGCTTC	MTCAACWTTT	CCCGGCTTCT	CCCGTTTTCR	AAYCTTNATC	CMTYCAAAAG	600
TTTAMTTTC	CCAAANRAAT	CCGYTTGCCA	CGTTGGCCGS	CGCTGGTTTT	CGMWCCTTRR	660
AMATCCNCCG	CGCGGSAAN	AMTTSGGNTT	CGSUCGGTCC	CCCGNAATAT	KCTTGGNGCT	720
ENAAATTGSS	GGGATCCGCG	CGGNAYCCCG	CTNCKGJSH	TCNCCAGTTT	GWACAAATTC	780
CTCCCTTCCA	AACTCCGGNC	CGGGGGGTGG	CGGCTNTTTC	CTMYNNAAA	AAGKGTTCGN	840
TTTTTTTGGG	INPAANTTCA	CGGCKNKTNT	CGNCHAACT	CTYCAANTTC	CANACCTTTA	900
ASAAAANCYV	YKTNLGGCG	TTTTTCCGCG	CANCTGTCGM	NMSSKNCGCG	AAAAAAAAGNK	960
CTNGGCTTAN	CNSNKTCTTT	TNKTWCCGCG	NMWNNSNMCI	NCBKKCNKPY	NGNSNMNCCT	1020
MYVYKCTRRN	SHNNNNKCCN	GSNCGGMYV	TMNNGNGMYE	NGNKSNNCCG	MSC	1073

(2) INFORMATION FOR SEQ ID NO:333:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1461 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:333:

GNSNGNKNTN	TMCAVCWYCT	SCACSGGGTC	TATTGCGGGC	GCAATYINGT	CKASAGATCT	60
CGATYTCGGC	AMNANAARTG	TGCTCGTCAA	TTTCAGKKTG	GTCKTCAAAY	GGGCCAGGCC	120
GNGACCRACA	CCCTGNGTCA	CCCCAAAAC	CAACAGCWTC	AAATWTCAAG	GCCRAGGCSC	180
TRTCAATYCC	CRASCAKTTA	ACCGTKTCW	TCRAAGGTGC	CRAACCAGGC	ACCCAGYTCA	240
CCGCCSGGCA	AWTCGCGCTG	CCGGCCGGTN	TCAGCCTGAT	TYCTGACCCT	RWTCTGTSGG	300
TGGYCAMCNT	GGTGAAGGCC	CWWCCGCCNA	AGAACTGGAG	GGCRAATTCC	CAGGANCCNA	360
GRAACCCNAG	GAACCCGCGG	TAKAANCCGG	CRAAACCRAG	GCCGYTGGCN	ATTCCNATTA	420
NAMSGGTTTG	CRACNTGGCC	RAACCGTTTT	CTTGGTCGGC	CTCGGCAACC	CTGGACCANT	480
TACCCCKTNC	CCGGNMCMAC	CYCGGGTNCT	TGKYCCCAAT	NTGCTCCCGC	GNRANTNGGC	540
CNAATTCCAG	GGCNCANCT	TTCCGGCCCN	AATTCCCYTG	GTTAATCACC	GGGCNCCCT	600
GGTTTTGGGC	AACCCNCYS	CTMTTTTAAA	CATTCCGSCC	CAAAATGGGNC	STTGGSAAT	660
TCTNTYCGGT	GGGGCSGGCR	ANMYTTCTCT	YCCCNAAASAN	CTTAMYCCAN	TTCCSSNTCC	720
CGGKCAAAWS	NGGGGGGGNA	AAGGGCCCCC	CGGNTSCKCC	GGGGKKGCC	CYGGKTTCAA	780
AANTTTCSGG	CKTSTMSCCS	NUTCSCCCC	CSGCCAAGRA	CCNGGGTTTT	TTTTGAACC	840
KCMANTCSSA	AMCCGCCSSC	CCMAAACCS	GCCTHAAWGR	RAYTTNKSCC	CTNAAACSGG	900
CCCCCAKVTY	SGGKTTTCNNC	CNCCSGKKGT	CCMTSTTTMM	MRCCCTTTGN	GNKTTTTTAN	960
MGSCCTTNNC	CACCCCYCK	GGKCSMNNA	GAATMYWKC	CNGGGGNAN	RSCCCSSTNN	1020
GGGKGGGGKG	MGAGYSCKT	CTKCGNCNN	YKNTTTCCCC	C		1081

(X2) INFORMATION FOR SEQ ID NO:334:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 986 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(E) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:334:

GNSNGRINKWN	ATMCAYCWYY	WTSCACCCGG	GMTCHATTGC	GGCCGCWKY	WNGTMAASAG	60
ATGTMGAAYT	CGGCACANAG	CGGCACAGAG	TGTGTGCATC	TGTGTGANAG	WTGTCAACCC	120
GTGCGCGCGC	GTGGTRASCA	CMCATTGCGR	AACACCAAC	CCGTCCGCGG	GYCACC3GCK	180
CCGCCTGCAA	AAYCCTCCAG	CCACCCVCRA	AACAAYWYCT	CGTGCAACSC	ARSCCGTTYC	240
CCGGCCGRAT	CCTGGKYCAS	WTGCGKTGC	CGTCCGCCAA	GGTACTGGCC	GWYCRANACC	300
CTYCGGGRA	ACCHAACTA	AATCTTTCNN	AATTTGNTT	CGCGTSCCC	CTTATMAATT	360
TTTTAAACCA	CCCAACCTY	CGCGCTTCT	CTCTTCCRA	WTGCTPWTY	WATNYCCGCA	420
CTGCTYNTK	KYCTYCHYCC	CTWCTAAAT	CTTCTATTC	TATATTCTCC	CTAAATGCCA	480
CTGCTGCTY	TGCTATNTCT	CTTCTTAAA	CTWAMANCAG	NGGTTTTCTY	CTTCCNAAAC	540
CTGTTGGCT	CAACCCNAAA	AATTAATATA	ATAATGGTGC	TNTCAAACCC	CGNCCCATY	600
CTATCGKCC	AMCCCGCGN	CTCTANFKC	CTAATCTTMM	AACCCCAAGT	CTATAASNTTG	660
CTAANAACCA	CTNCCGCTCA	CTAANAACANY	CTTNTTCCNY	CTNTTCCGMN	CTATGGCTNN	720
CTAANAACCA	AATACTNTY	CTYCTAATAA	AAMMMSGGYC	CTMCCCGAAA	CTTTTTTTGN	780
CTAANAACCA	AAGCTTTTT	CTAAGCCDAN	CTYCTCTNCC	CTCTMANTGG	CTNSGGARTKT	840
CTAANAACCA	CTGKYCCMAA	CTGNGGPRANA	CTAPCCCCAA	CTCTNNNTN	CTNNGCCNST	900
CTAANAACCA	CTNTYCTMAA	CTAAGCTTCC	CTNTTCCAA	CTAAMCCGNN	CTAAGARTCTN	960
CTAANAACCA	CTNTYCTTCA	CTMMM				1020

- (A) LENGTH: 1074 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

```

NGNGGGNKRN  ATMMAYCWCT  SATYYACCSN  GGMNMWATTG  CGGCCRMWAT  CTNGTMKASA      60
GATCTMGAAG  YTCGGCAAAG  AGYATKCTCG  GGGGCCAGAT  TTNTGGCCCC  CAACCGCCCGC    120
ACTTTGCAYW  TCAACAKTCC  SGGTGCCCCA  AAAAAWTCWT  ACCCCCATMC  TYCKTGCASM     180
ASYTGCGCCC  RATTRAACAC  CCGGCCGGCW  TGCTGCGCCA  GGTATTYCAS  CAGYTCAAAY     240
YCTTTKTAGK  TAAAAATCAG  CCGGCCGGCA  CACAGCCGGG  CGGTKTAGGT  GCCTYCRTCA     300
ATMACCAGCY  CGCCCAAGGY  CACCTTGCCC  AAAAYCTCCT  GGGTCAGCCA  AATTYCCGCS     360
CGCGCCAACM  ACCANCCCCA  TYCTGGCCTC  AATCYCACCG  CGCCCGGTGY  TAAAMMANMA     420
BRATCTCKTC  MANCCCCCAN  TCAGCSYTN  CNGCMACAGC  CGCCCTTCTT  CAMACCGCCA     480
ATACCGCGWT  CAACCGCCCG  CTCAAACCTA  ACAGGCGGNC  AGGCCTCCCC  CGGANSAAAG     540
CTCTTACSCC  NNYYAANAAA  MAAGNTCTGT  TTTCCCGCTC  TASAASNAAA  AANCCCCSGC     600
CGGGCCTTCN  NMMGGGTTTC  GGGMANANAA  AARCNCCGGN  GGAACGNATC  CGAAAMCTCC     660
CAAGTCNCMT  TWAWAACYCN  NNAACCCCCC  ANTTTTGGGA  AAGGNTCCCC  NTTMYCCCCC     720
TTTTASGKTS  GGGMMYYCTY  TAAAAAATT  CCCCCAAAAG  CCGCGGGAAG  GGTGMAMCTG     780
GGNAAATTC  CAAMCCNWC  TTTTYNGGT  TMCGGGGGRA  AATTYCNCCT  CCYYNNNGGG     840
CSSGSNNNAT  TAYGGSNMT  TTTNNAWTM  NSGKKTSAAM  YNNKCCMNNN  SNNMSMANNK     900
TWAMCKCCCN  CCTCNGNGKY  CCGYNQCCG  GNAGNGGRAS  MKCCNANMAA  AYASGNTTNK     960
CGGAAMMCNN  AATKGNNSC  CCGGASMCN  NNMMAATMT  CNCNKCNSNN  AANRGMRACN    1020
CCNSNSGMN  RRGAAARMTN  YCCCCCGSKM  SKGNKAAAAW  GKYCCCCCCM  AAAG           1074
  
```

10 INFORMATION FOR SEQ ID NO:336

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1198 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

12. SEQUENCE DESCRIPTION: SEQ ID NO:336

```

GNGNENHNT  MTAAATCTTT  TTACGCTG  NTGMAWTTT  GGTGGAWWT  TGTGASAG      60
TGTGAAAT  TGGGAGAG  AATGCTG  TGTGCTG  TGTGCTG  TGTGCTG      120
ATTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      180
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      240
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      300
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      360
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      420
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      480
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      540
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      600
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      660
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      720
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      780
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      840
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      900
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG      960
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG    1020
TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG  TGTGCTG    1074
  
```

TATKSAGMGG	TKCCGMAGMK	CCSCSTTTT	TKTGANAAMN	MSMRKNKKTG	CGMGYTCTSC	960
GGGNTTTTGT	GAGTAKTGGG	CCSSSMWGAC	WCSGMCMGNG	AGKNKTNNTS	YANTGARCGY	1020
MNNSKTMKMT	MSCSCGCGNA	GGAGNGCCCC	CSANGMSTGY	NKGGNMSSNG	ARAKGATGGS	1080
GGCCNCGMNN	MGMGGANMGA	SANNGMGMR	GGGGGKTGKC	TCKCSCCGNS	CSANGRAGAA	1140
GKTCNGSCGC	CGMGKYGKT	KTKTKNKTGG	YSTCMSSMMM	NAGAAAAGAG	AGG3C	1195

(2) INFORMATION FOR SEQ ID NO:337:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3572 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

CCATCTGATC	CTTGCCCAACC	AGCATGCGAG	TGGGAACGAT	CCCTTCATTC	AGCATTTCGA	60
TGGTTTCTTG	AAAACCGGAC	ATGGCACTCC	AGTCCGCTTC	CCGTTCCGCT	ATCGGCTGAA	120
TTTGATTCCG	AGTGAGATAT	TTATGCCAGC	CAGCCACAGC	CAGACCGCGC	GAGACAGAAC	180
TTAATGCGCC	CGCTAACAGC	CCGATTTCCT	GGTGACCCAA	TGCGACCAGA	TGCTCCACGC	240
CCAGTCCCGT	ACCGTCTTCA	TGGGAGAAAA	TAATACTGTT	GATGGGTGTC	TGGTCAGAGA	300
CATCAAGAAA	TAACGCCCGA	ACATTAGTGC	AGGAGAGTTC	CACAGCAATC	GCATCCTGGT	360
CATCCAGCGG	ATAGTTAATG	ATCAGCCGAC	TGACGCGTTG	CCCGAGAAGA	TTGTGCACCG	420
CCGCTTTACA	GGCTTCGAGC	CCGCTTCGTT	CTAGCATCGA	CACCACGAGC	CTGGCACCCA	480
CTTGATCGGC	GCGAGATTTA	ATCGCCGCGA	CAATTTGCGA	CGGCGCGTGC	AGGGCCAGAC	540
TGGAGGTGGC	AACGCCAATC	AGCAACGAGT	CTTTGCCCGC	CAGTTGTTGT	GCCACGCGGT	600
TGGGAATGTA	ATTGAGCTTC	GCATCGCCG	CTTCCACTTT	TTCCCGCGTT	TTCCGAGAAA	660
CTGCTCTGGC	CTGCTTCAGC	AAGCCGGA	CGGTCTGATA	AGAGACACCG	GCATACTCTG	720
CGACATCGTA	TAACCTTACT	GGTTTCACAT	TCACCACTCT	GAATTGACTC	TCTTCCGGGC	780
CTATCATGTC	CATACCCCGA	AAGGTTTTGC	CGATTGAT	GTGTCCCGCG	ATCTGACGCG	840
CTCTCTTTAT	CGGACTCTTC	CATTAGGAAG	CAGCCCACTA	GTAGCTTACG	CCGCTTGAGC	900
AAGCCCGCGC	CAAGGAATCG	TGCATGCAAG	CAGATGCGCG	CCAACTCTCC	CCGCGCCAGC	960
GGGCTTCCGA	CGATACCGAG	CGCAAAACAA	CGCTCATGCA	CGCCGAACTC	CGGAGCCCGA	1020
TCTTCCCGAT	CGGTGATCTC	CGCGATATAG	CGCCGAGCAA	CGCCAGCTCT	CGCGCGCGTC	1080
ATCCCGGCGA	CGATGCGCTC	CGGCTAGAGG	ATCGAGATCT	CGATCCCGCG	AAATTAATAG	1140
TACTCACTAT	AGGGGAATTC	CGAGCGGATA	ACAATTCGCC	TGTAGAAATA	ATTTTGTGTA	1200
ACTTTAAGAA	CGAGATATAG	ATATGCGGCA	TCATCATCAT	CATCAGCTGA	TCCATATCAT	1260
TGGGACGAGC	CGACATCTCT	CGCAACAGAG	CGGCTGAGAG	CGGCTGAGAG	CGGCTGAGAG	1320
TAGCTCTGAT	CACATCGCGC	CGCTTCTGAT	CATTGAGGAG	GAGATGCGCG	TGGACAGCGC	1380
CGGCAAGATC	AGGTAGCTTA	CGAACTTCTA	CGCTTCTGAT	GAGATGAGCT	CGGCGGAGCG	1440
AGCGCGCTCG	AAAGGAGCGA	AGGCTTCTCT	CGAACTTCTA	CGCTTCTGAT	CGGCGGAGCG	1500
AGTACGCTCG	CGCTCTCTCT	CGCTTCTCTA	CGCTTCTCTA	CGCTTCTCTA	CGCTTCTCTA	1560
CGTCTTCTAC	CTCTCGGCTC	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1620
TGAGGCGAGC	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1680
CGGCTTCTCT	ATCTCTCTCT	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1740
CAAGCTGAA	CGAAAGCTCC	TGGCGGCTAT	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1800
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1860
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1920
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	1980
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2040
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2100
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2160
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2220
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2280
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2340
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2400
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2460
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2520
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2580
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2640
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2700
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2760
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2820
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2880
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	2940
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3000
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3060
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3120
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3180
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3240
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3300
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3360
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3420
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3480
CGGCTTCTCT	CGCTTCTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	CGGCTTCTA	3540

```

CGCGCGGGCT GGGTTGSCAT CGAAAACCCG GCGGAACCAG GCGATTTCGA TGATCGACGG 2280
GCCCCCCCCG GACGGGTACC CGATCATCAA CTACGAGTAC GCCATCGTCA ACAACCGGCA 2340
AAAGGACGCC GCGACGCGCG AGACCTTGCA GGCATTTCTG CACTGGGCGA TCACCGACCG 2400
CAACAAGGCC TCGTTGCTCG ACCAGGTTCA TTTCCAGCCG CTGCCGCCCG CGGTGGTGAA 2460
GTTGTCTGAC GCGTTGATCG CGACGATTCG CAGCGCTGAG ATGAAGACCG ATGCCGCTAC 2520
CCTCGCGCAG GAGGCAGGTA ATTTCCGAGCG GATCTCCGGC GACCTGAAAA CCCAGATCGA 2580
CCAGGTGGAG TCGACGGCAG GTTCGTTGCA GGGCCAGTGG CGCGGCGCGG CGGGGACGGC 2640
CGCCCAGGCC GCGGTGGTGC GCTTCCAAGA ACCAGCCAAT AAGCAGAAGC AGGAACTCGA 2700
CGAGATCTCG ACGAATATTC GTCAGGCGCG GTCCTCAATC TCGAGGGCCG ACGAGGAGCA 2760
GCAGCAGGCG CTGTCTCTCG AAATGGGCTT TGGATTGAGC TTCGCGCTGC CTGCTGGCTG 2820
GGTGGAGTCT GACGCCGCCC ACTTCGACTA CGGTTGAGCA CTCCTCAGCA AAACCACCGG 2880
GGACCCGCCA TTTCCCGGAC AGCCGCGCGC GGTGGCCAAT GACACCCGTA TCGTGCTCGG 2940
CCGGCTAGAC CAAAAGCTTT ACGCCAGCGC CGAAGCCACC GACTCCAAGG CCGCGGCCCG 3000
GTTGGGCTCG GACATGGGTG AGTTCTATAT GCCCTACCCG GGCACCCGGA TCAACCAGGA 3060
AACCCTCTCG CTYGACGCCA ACGGGGTGTC TGGAAGCGCG TCGTATTACG AAGTCAAGTT 3120
CAGCGATCCG AGTAAGCCGA ACGGCCAGAT CTGGACGGGC GTAATCGGCT CGCCCGCGGC 3180
GAACGCACCG GACGCGCGGC CCGCTCAGCG GTGGGCGGAA TCGATTCCCG CTTTGGTCCG 3240
CAACCCGGTG GACAAGGGCG TCGCGGCGCG GCTGGCGGAA TCGATTCCCG CTTTGGTCCG 3300
CCCGCCGCGC GCGCGCGCGC GCGAAGTCCG TCCTACCCCG ACGACACCGA CACCGCAGCG 3360
GACCTTACCG GCGTGAAGAT TCTGCAGATA TCGATCACAC TGGCGGCCCG TCGAGCACCA 3420
CGACCCACCG CACTCAGATC CCGCTGCTAA TAAAGCCCGA AAGGAAGCTG ACTTGGCTCC 3480
TCCGACCGCT GAGCAATAAC TACCATACCG CTTTGGGCGC TGTAAACCGG TCTTCAGCGC 3540
TTTTTTGCTG AAAGGAGGAA CTATATCCCG AT
3572

```

(2) INFORMATION FOR SEQ ID NO:338:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2) MOLECULE TYPE: peptide

3) SEQUENCE DESCRIPTION: SEQ ID NO:338:

```

Val Glu Phe Glu Ser Gly Gly Asp Asn Ser Pro Ala Val Tyr Lys Lys
      5              10              15
Asp Gly Lys Arg
      20

```

4) INFORMATION FOR SEQ ID NO:339:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2) MOLECULE TYPE: peptide

3) SEQUENCE DESCRIPTION: SEQ ID NO:339:

1. *Chlorophyll a* and *Chlorophyll b* were determined by the method of Arar and Collins (1971) using a Shimadzu 1601 UV-Visible Spectrophotometer. The concentration of chlorophyll was expressed in $\mu\text{g mL}^{-1}$.

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

CTAGTTAGTA CTCAGTCGCA GACCGTG

27

(2) INFORMATION FOR SEQ ID NO:344:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

GCAGTGACGA ATTCACTTCC ACTCC

28

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2412 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

ATATATGGGCC	ATCATCATCA	ATATCATCTG	ATCGACATCA	TGCGGACGAG	CCCGACATCG	61
TGCGAACAGG	CGCGCGCGGA	CGCGGTCCAG	CGCGCGCGGG	ATAGCCTCGA	TCACATCCGC	120
CTCGCTCGGG	TCATTGAGCA	CGACATCGCC	CTCGACAGCG	CGCGCAAGAT	CACCTACCGC	180
ATCAAGCTCG	AAGTGTCTTT	TAAGATGAGC	CGCGCGCAAC	CGAGGGGGTC	GAACACCACG	240
ACCGCTTCGC	CTGAAACCGG	CGCGCGCGCC	GGTACTGTCT	CGACTACCGC	CGCGTCTCGC	300
CGCGTCACTT	TGCGCGAGAC	CGGTAGCAGC	CTGCTCTAGC	CGCTGTTCAA	CGTGTGGGGT	360
CGCGCGCTTC	ACGAGAGGTA	TGCGAAGCTG	AGGATCAGCG	CTCAGGGGCA	CGGTTCCTGT	420
CGCGCGGATC	CGCAGCGCGT	CGCGCGGAGC	CTCAACATTC	CGCGCTTCGA	CGGTATCTCT	480
CGCGAAGGTC	ATATCGCGCG	CGCGAAGGCG	CTCATGAACA	TCGCGCTAGT	ATATCTCGCT	540
CGCGAAGGTC	ACTACACCTT	CGCGGAGGTC	ACCGAGCAGT	TCAGCTGAA	CGGAAAGTTC	600
CGCGCGGCGA	CTATACAGCG	CGCGATCAAA	ACCTGGGAGC	ACCGCGAGAT	CGGTGCGCTC	660
ACCGCGCGCG	TGAAGCTTCC	CGCGACCGCG	CTAGTTCGCG	TGCACCGCTT	CGACGGGTCC	720
CTTACACCTT	TCTTCTTCA	CGAGTACCTT	TCGAAGCAAG	ATCGCGAGGT	CTCGGCGAAG	780
CGCGCGCGCT	TCGCGACCA	CGTGGACTTC	CGCGCGGTTC	CGGTGCGCTT	CGGTGAGAAC	840
CGCAACCGCG	CGATGCTGAC	CGGTTCGCGT	GAGACACCGG	GCTGCGTGGC	CTATATCGGC	900
ATCAGCTTCC	TGCGACAGCG	CACTCAAGCG	CGACTCGCGC	AGCGCGAAGT	AGGCAATAGC	960
CTTGGCAATT	TCTGCTTTCG	CGACGCGTAA	AGCATTCAGC	CGCGCGCGCT	TGGTTTGGCA	1020
CGGAAACCGT	CGCGCAACCA	CGCGATTCCT	ATGATCGAGC	CGCGCGCGCT	CGATTTGTA	1080
CGATATATCA	ATTAACACTT	CGCGATTCCT				1140

```

AATTTTCGAGC GGATCTCCGG CCACCTGAAA ACCCAGATCG ACCAGGTGGA GTCGACGGCA 1380
GGTTTCGTTGC AGGGCCAGTG GCGCGGCGCG GCGGGGACGG CCGCCCAGGC CGCGGTGGTG 1440
CGCTTCCAAG AAGCAGCCAA TAAGCAGAAG CAGGAAC TCG ACGAGATCTC GACGAATATT 1500
CGTCAGGCGC GCGTCCAATA CTCGAGGGCC GACGAGGAGC AGCAGCAGGC GCTGTCTCTG 1560
CAAATGGGCT TTGTGCCAC AACGGCCGCC TCGCCCGCGT CGACCGCTGC AGCGCCACCC 1620
GCACCGGCGA CACCTGTTGC CCCCCACCA CCGGCGCGCG CCAACACGCC GAATGCCCCAG 1680
CCGGGCGATC CCAACGCAGC ACCTCCGCGG GCCGACCCGA ACGCACC GCCACCTGTC 1740
ATTGCCCCAA ACGCACCACA ACCTGTCCGG ATCGACAACC CGGTGGGAGG ATTCAGCTTC 1800
GCGCTGCCTG CTGGCTGGGT GGAGTCTGAC GCCGCCACT TCGACTACGG TTCAGCACTC 1860
CTCAGCAAAA CCACCGGGA CCGCCATTT CCGGACAGC CGCCGCCCGT GGCCAATGAC 1920
ACCCGTATCG TGCTCGGCG GCTAGACCAA AAGCTTTACG CCAGCGCCGA AGCCACCGAC 1980
TCCAAGGCCG CGGCCGGTT GGGCTCGGAC ATGGGTGAGT TCTATATGCC CTACCCGGGC 2040
ACCCGGATCA ACCAGGAAC CGTCTCGCTC GACGCCAACG GGGTGTCTGG AAGCGCGTCG 2100
TATTACGAAG TCAAGTTCAG CGATCCGAGT AAGCCGAACG GCCAGATCTG GACGGGCGTA 2160
ATCGGCTCGC CCGCGGCGAA CGCACCAGC GCGGGGCCCC CTCAGCGCTG GTTTGTGGTA 2220
TGGCTCGGGA CCGCCAACAA CCGGTGGAC AAGGGCGCGG CCAAGGCGCT GGCCGAATCG 2280
ATCCGGCCTT TGGTCGCCCC GCGCGCGCGC CCGGCACCGG CTCCTGAGA GCGCGCTCCG 2340
CGCGCGCGCG CCGCGCGGGA AGTGGCTCCT ACCCGACGA CACCGACAC CCAGCGGACC 2400
TTACCGGCCT GA
2410

```

(2) INFORMATION FOR SEQ ID NO:346:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 902 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: protein

(3) SEQUENCE DESCRIPTION: SEQ ID NO:346:

```

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser
1      5      10
Pro Thr Ser Trp His Gln Ala Ala Ala His Ala Val Gln Arg Ala Arg
15      20      25      30
Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
35      40      45
Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu His Val
50      55      60
Ser Phe Lys Met Arg Phe Ala Gln His Arg Gly Ser Lys His Ser Ser
65      70      75      80
Gly Ser Phe His Thr Gly Ala His Ala Gly Thr Val Ala Thr Thr Pro
85      90      95
Ala Ser Ser Phe Val Thr Leu Ala His Thr Gly Ser Thr Leu Leu Tyr
100      105      110
Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
115      120      125
Gly Thr Ile Thr Ala Gln His Thr Gly Ser Gly Ala Gly Ile Ala Gln
130      135      140
Ala Ala Ala Gly Thr Thr Asp Thr His Ala Ser Asp His Thr Thr Ser
145

```


Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
 625 630 635 640
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685
 Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys
 690 695 700
 Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile
 705 710 715 720
 Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp
 725 730 735
 Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala
 740 745 750
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro
 755 760 765
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala
 770 775 780
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu
 785 790 795 800
 Pro Ala

2) INFORMATION FOR SEQ ID NO:347:

(1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 34 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

1) MOLECULE TYPE: Other

X1 SEQUENCE DESCRIPTION: SEQ ID NO:347:

GGATCCAAAC CACCGAGCGG TTCGCTGAA ACCG

34

2) INFORMATION FOR SEQ ID NO:348:

(1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 37 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

1) MOLECULE TYPE: Other

X1 SEQUENCE DESCRIPTION: SEQ ID NO:348:

GGATCCAAAC CACCGAGCGG TTCGCTGAA ACCG

(1) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1962 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:349:

```

CATATGGGGCC ATCATCATCA TCATCAGGGA TCCAAACCAC CGAGCGGTTT CCCTGAAACG      60
GGGCGCGGCG CCGGTACTGT CGCGACTACC CCCGCGTCGT CCGCGGTGAC GTTGGCGGAG      120
ACCGGTAGCA CGCTGCTCTA CCGCTGTGTC AACCTGTGGG GTCCGGCCTT TCACGAGAGG      180
TATCCGAACG TCACGATCAC CGCTCAGGGC ACCCGTTCTG GTGCCCGGAT CCGCCAGGSC      240
GGCGCGGGGA CGGTCAACAT TGGGGCGCTCC GACGCCTATC TGTCCGAAGG TGATATGGCC      300
GGGCACAAGG GGCTGATGAA CATCGCGCTA GGCATCTCCG CTCAGCAGGT CAACTACAAC      360
CTGCGCGGAG TGAGCGAGCA CTTCAAGCTG AACGMAAAG TCCTGGCGGC CATGTACCAG      420
GGCACCATCA AAAGCTTGGA TTAGCTGAGG ATCGCTGCGC TCAACCGCGG CTTGAAGCTG      480
CGCGCGCAGG CGGTAGTTCC CTTGCAAGCG TCCGACGGGT CCGGTGACAC CTTCTTCTTC      540
ACCGAGTAGC TGTCAAGCA AGATCGGAG GCTGCGGCA AGTCCGCGCG CTTCCGCCAC      600
ACCGTTCGACT TCCCGCGCGT GCGCGGTGCG CTGCGTGAGA ACCGCAACGG CGGCTGCTTC      660
ACCGGTTGCG CCGAGAGAGC GCGGTGCGT GCTATATGCG GCATCAGCTT CCGCGACGAG      720
CGCAGTCAAC GGGGACTCGG CGAGGCGCAA CTAGGCAATA GCTCTGGCAA TTTCTTGTTC      780
CGCGCGATTT CGATCATCGA CCGCGCGCGC CGCGAUGGTT ACCCGATCAT CAACTACGAG      840
TACCGCATCG TCAACAACCG GMAAAGGAC GCGCGCACCG CGCAGACCTT GCAGGCATTT      900
CTGCACTGCG CGATCACCGA CCGCAACAG GCTTGTTC TCGACGAGGT TCATTTCCAG      1020
CGCGTGCCTC CCGCGCTGCT GAAGTGTCTT GACCGCTTGA TCGCGACGAT TTCCTCGGGA      1080
GCTGCGCAJG GCGGAGGCTC AGGTGGAAGT TGTGCGCGGA GCTGCGCCAC AACCGCGCGC      1140
TCCCGCGCGT CGACCGCTTC AGCGCGACCG CGACCGCGGA GACCTGTTGC CCGCGCACCA      1200
CGCGCGCGTG CCAACACCGC GAATGCGGAG TCGCGCGATC CGACCGCAGC ACCTCGCGCG      1260
GCTGACCGGA ACCGACCGCG CGCACCTGTC ATTGCGCGAA ACCGACCGCA ACCTGTCGCG      1320
ATCGACCAAC CGGTGCGAGG ATTCAGCTTC GCGCTGCTG CTGCTGCGGT TGAATGTCAG      1380
CGCGCGCAAT TCGACTACCG TTCAGCACTC TCGAGCAAAA TCGCGCGGGA TCGCGCATTT      1440
TCCCGGAGAC CCGCGCGCGT GCGCAATGAC ACCGCTATCG TCGTGGCGCG GCTAGACCAA      1500
AAGCTTTACG CGAGCGCGCA AGCCACCGAC TCGAAGGCTG CCGCGCGGTT GCGCTCGGAC      1560
ATCGGTGAGT TGTATATGCG CTACCGCGCG ACCCGGATCA ACCAGGAAAC CGTCTCGCTC      1620
TACCGCAACG CGGTGCTCTG AAGCGCTGCG CATTACGAAG TCAAGTTTCA GATCGGAGT      1680
AAGCGGAACG CGCAGATCTC GACCGCGCTA ATCGGCTGCG TCGCGCGGAA CGCACCGGAC      1740
CGCGCGCGCG CTCAGCGCTC TTTCTGCTA TCGCTCGGA TCGCTAAGAA TCGCGCTGAA      1800
AAGCGCGCGG CGAAGCGGCT TTTGAAATCG ATCGCGCTT TTTTGGCGCG TCGCGCTGAA      1860
TGGGCGCGG CTCTCGGA TTTGCTGCG TCGCGCGCT TCGCGCGGAA TCGCGCTGAA      1920
TGGGCGCGG TTTGCTGCG TTTGCTGCG TCGCGCGCT TCGCGCGGAA TCGCGCTGAA      1980

```

2. INFORMATION FOR SEQ ID NO:349:

(1) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 612 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

Met Gly His His His His His His Gly Ser Lys Pro Pro Ser Gly Ser
 1 5 10 15
 Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
 20 25 30
 Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr Pro Leu
 35 40 45
 Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr
 50 55 60
 Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
 65 70 75 80
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly
 85 90 95
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
 100 105 110
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
 115 120 125
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 130 135 140
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 145 150 155 160
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 165 170 175
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 180 185 190
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 195 200 205
 Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 210 215 220
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 225 230 235 240
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 245 250 255
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 260 265 270
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
 275 280 285
 Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Gln Tyr Ala Ile Val Asn
 290 295 300
 Asn Arg Gln Lys Asp Ala Ala Thr Ala His Thr Leu Gln Ala Phe Leu
 305 310 315 320
 Asp Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asn Thr Val
 325 330 335
 Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
 340 345 350
 Ile Ala Thr Ile Ser Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly
 355 360 365
 His Ser Gly Gly Ser Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
 370 375 380
 Ala Ala Ala Thr Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
 385 390 395 400

Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser
435 440 445
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp
450 455 460
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
465 470 475 480
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
485 490 495
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
500 505 510
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
515 520 525
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
530 535 540
Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
545 550 555 560
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
565 570 575
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
580 585 590
Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
595 600 605
Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
610 615 620
Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr
625 630 635 640
Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala
645 650

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a) and Rule 39)

Applicant's or agent's file reference 9532-023-228	IMPORTANT DECLARATION	Date of mailing (day/month/year) 22 JUNE 1999
International application No. PCI/US99/03265	International filing date (day/month/year) 17 FEBRUARY 1999	(Earliest) Priority Date (day/month/year) 18 FEBRUARY 1998
International Patent Classification (IPC) or both national classification and IPC Please See Continuation Sheet.		
Applicant CORIXA CORPORATION		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below

1. ☐ The subject matter of the international application relates to:
 - a. ☐ scientific theories.
 - b. ☐ mathematical theories.
 - c. ☐ plant varieties.
 - d. ☐ animal varieties.
 - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
 - f. ☐ schemes, rules or methods of doing business.
 - g. ☐ schemes, rules or methods of performing purely mental acts.
 - h. ☐ schemes, rules or methods of playing games.
 - i. ☐ methods for treatment of the human body by surgery or therapy.
 - j. ☐ methods for treatment of the animal body by surgery or therapy.
 - k. ☐ diagnostic methods practiced on the human or animal body.
 - l. ☐ mere presentations of information.
 - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:

☐ the description
☐ the claims
☐ the drawings
3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the prescribed requirements prevents a meaningful search from being carried out.

☒ it does not comply with the prescribed standard

☐ it is not in the prescribed machine-readable form
4. ☐ Other comments

**DECLARATION OF NON-ESTABLISHMENT OF
INTERNATIONAL SEARCH REPORT**

International application No.
PCT/US99/03265

The International Patent Classification (IPC) or National Classification and IPC are as listed below

IPC(5): A61K 38/00; C07K 1/00; C07K 16/00; C12Q 1/68; C12P 19/34; C07H 21/02, 21/04
US Cl. 530/300, 350, 387.1; 435/6, 91.1, 91.2; 536/23.1, 24.3, 24.31, 24.32, 24.33